

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
4 January 2001 (04.01.2001)

PCT

(10) International Publication Number
WO 01/00672 A1

(51) International Patent Classification⁷: C07K 14/47, C07H 21/04, C12N 15/63, 1/21, C12P 21/02

(74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(21) International Application Number: PCT/US00/18184

(22) International Filing Date: 29 June 2000 (29.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/342,687 29 June 1999 (29.06.1999) US

(71) Applicant: MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors: HOLTZMAN, Douglas, A.; 821 Centre Street, #6, Jamaica Plain, MA 01230 (US). BARNES, Thomas, M.; 22 Hanson Street, #2, Boston, MA 02118 (US). FRASER, Christopher, C.; 52 Grassland Street, Lexington, MA 02421 (US). SHARP, John, D.; 245 Park Avenue, Arlington, MA 02476 (US).

Published:

— With international search report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SECRETED PROTEINS AND USES THEREOF

(57) Abstract: The invention provides isolated nucleic acid molecules, designated TANGO 244, TANGO 246, TANGO 275, TANGO 300, and MANGO 245 which encode wholly secreted or membrane-associated proteins. The invention also provides antisense nucleic acid molecules, expression vectors containing the nucleic acid molecules of the invention, host cells into which the expression vectors have been introduced, and non-human transgenic animals in which a nucleic acid molecule of the invention has been introduced or disrupted. The invention still further provides isolated polypeptides, fusion polypeptides, antigenic peptides and antibodies. Diagnostic, screening and therapeutic methods utilizing compositions of the invention are also provided.

WO 01/00672 A1

SECRETED PROTEINS AND USES THEREOF

This application claims priority to co-pending U.S. Application No. 09/342,687, filed June 29, 1999, the entire contents of which are incorporated herein by reference in its entirety.

5

Background of the Invention

Many secreted proteins, for example, cytokines, play a vital role in the regulation of cell growth, cell differentiation, and a variety of specific cellular responses. A number of medically useful proteins, including erythropoietin, 10 granulocyte-macrophage colony stimulating factor, human growth hormone, and various interleukins, are secreted proteins. Thus, an important goal in the design and development of new therapies is the identification and characterization of membrane-associated and secreted proteins and the genes which encode them.

Many membrane-associated proteins are receptors which bind a ligand and 15 transduce an intracellular signal, leading to a variety of cellular responses. The identification and characterization of such a receptor enables one to identify both the ligands which bind to the receptor and the intracellular molecules and signal transduction pathways associated with the receptor, permitting one to identify or design modulators of receptor activity, *e.g.*, receptor agonists or antagonists and 20 modulators of signal transduction.

Summary of the Invention

The present invention is based, at least in part, on the discovery of cDNA molecules encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300, and 25 MANGO 245, all of which are predicted to be either wholly secreted or transmembrane proteins. These proteins, fragments, derivatives, and variants thereof are collectively referred to as a "polypeptides of the invention" or "proteins of the invention." Nucleic acid molecules encoding the polypeptides or proteins of the invention are collectively referred to as "nucleic acids of the invention."

30 The nucleic acids and polypeptides of the present invention are useful as modulating agents in regulating a variety of cellular processes. Accordingly, in one

aspect, this invention provides isolated nucleic acid molecules encoding a polypeptide of the invention or a biologically active portion thereof. The present invention also provides nucleic acid molecules which are suitable for use as primers or hybridization probes for the detection of nucleic acids encoding a polypeptide of the invention.

5 The invention features nucleic acid molecules which are at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98%) identical to the nucleotide sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, the nucleotide sequence of the cDNA insert of a clone deposited with ATCC® as Accession Number 207220 (the "cDNA of ATCC® Accession Number 207220"), the nucleotide sequence of the
10 cDNA insert of a clone deposited with ATCC® as Accession Number 207223 (the "cDNA of ATCC® Accession Number 207223"), the nucleotide sequence of the cDNA insert of a clone deposited with ATCC® as Accession Number PTA-248 (the "cDNA of ATCC® Accession Number PTA-248"), or the nucleotide sequence of the cDNA insert of a clone deposited with ATCC® as Accession Number PTA-293 (the
15 "cDNA of ATCC® Accession Number PTA-293").

 The invention features nucleic acid molecules which include a fragment of at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2400, 2600, 2800, 3000, 3200, 3400, 3600, 3800, or 4000) nucleotides of the nucleotide sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12,
20 13, 15, 16, 18, 19, 22, 24, or 25, the nucleotide sequence of the cDNA of ATCC® Accession Number 207220, the nucleotide sequence of the cDNA of ATCC® Accession Number 207223, the nucleotide sequence of the cDNA of ATCC® Accession Number PTA-248, or the nucleotide sequence of the cDNA of ATCC® Accession Number PTA-293, or a complement thereof.

25 The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98%) identical to the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, the amino acid sequence encoded by the cDNA of ATCC® Accession Number 207220, the amino acid sequence encoded
30 by the cDNA of ATCC® Accession Number 207223, the amino acid sequence

encoded by the cDNA of ATCC[®] Accession Number PTA-248, or the amino acid sequence encoded by the cDNA of ATCC[®] Accession Number PTA-293.

In preferred embodiments, the nucleic acid molecules have the nucleotide sequence of SEQ ID NOs: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, the
5 nucleotide sequence of the cDNA of ATCC[®] Accession Number 207220, the nucleotide sequence of the cDNA of ATCC[®] Accession Number 207223, the nucleotide sequence of the cDNA of ATCC[®] Accession Number PTA-248, or the nucleotide sequence of the cDNA of ATCC[®] Accession Number PTA-293, or a complement thereof.

10 Also within the invention are nucleic acid molecules which encode a fragment of a polypeptide having the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, or a fragment including at least 15 (25, 30, 50, 100, 150, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, or 1400) contiguous amino acids of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, the amino acid sequence
15 encoded by the cDNA of ATCC[®] Accession Number 207220, the amino acid sequence encoded by the cDNA of ATCC[®] Accession Number 207223, the amino acid sequence encoded by the cDNA of ATCC[®] Accession Number PTA-248, or the amino acid sequence encoded by the cDNA of ATCC[®] Accession Number PTA-293.

The invention includes nucleic acid molecules which encode a naturally
20 occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, the amino acid sequence encoded by the cDNA of ATCC[®] Accession Number 207220, the amino acid sequence encoded by the cDNA of ATCC[®] Accession Number 207223, the amino acid sequence encoded by the cDNA of ATCC[®] Accession Number PTA-248, or the amino acid sequence
25 encoded by the cDNA of ATCC[®] Accession Number PTA-293, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule consisting of a nucleic acid sequence encoding SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, the nucleotide sequence of the cDNA of ATCC[®] Accession Number 207220, the nucleotide sequence of the cDNA of ATCC[®] Accession Number 207223, the nucleotide
30 sequence of the cDNA of ATCC[®] Accession Number PTA-248, or the nucleotide

sequence of the cDNA of ATCC[®] Accession Number PTA-293, or a complement thereof under stringent conditions.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 60%, preferably 65%, 75%, 85%, 95%, or
5 98% identical to the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, the amino acid sequence encoded by the cDNA of ATCC[®] Accession Number 207220, the amino acid sequence encoded by the cDNA of ATCC[®] Accession Number 207223, the amino acid sequence encoded by the cDNA of ATCC[®] Accession Number PTA-248, or the amino acid sequence encoded by the cDNA of
10 ATCC[®] Accession Number PTA-293.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 60%, preferably 65%, 75%, 85%, or 95% identical the nucleic acid sequence encoding SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, and isolated polypeptides or proteins
15 which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NOs: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, or complement thereof, the non-coding strand of the cDNA of ATCC[®] Accession Number 207220, the non-coding strand of the cDNA of ATCC[®] Accession
20 Number 207223, the non-coding strand of the cDNA of ATCC[®] Accession Number PTA-248, or the non-coding strand of the cDNA of ATCC[®] Accession Number PTA-293.

Also within the invention are polypeptides which are naturally occurring allelic variants of a polypeptide that includes the amino acid sequence of SEQ ID
25 NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, the amino acid sequence encoded by the cDNA of ATCC[®] Accession Number 207220, the amino acid sequence encoded by the cDNA of ATCC[®] Accession Number 207223, the amino acid sequence encoded by the cDNA of ATCC[®] Accession Number PTA-248, or the amino acid sequence encoded by the cDNA of ATCC[®] Accession Number PTA-293, wherein the
30 polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule having the sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18,

19, 22, 24, or 25, or a complement thereof, under stringent conditions. Such allelic variant differ at 1%, 2%, 3%, 4%, or 5% of the amino acid residues.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NOs: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, the cDNA of ATCC[®] Accession Number 207220, the cDNA of ATCC[®] Accession Number 207223, the cDNA of ATCC[®] Accession Number PTA-248, or the cDNA of ATCC[®] Accession Number PTA-293, or a complement thereof. In other embodiments, the nucleic acid molecules are at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3200, 3400, 3600, 3800, 4000, or 4200) nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule consisting of the nucleotide sequence of SEQ ID NOs: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, the cDNA of ATCC[®] Accession Number 207220, the cDNA of ATCC[®] Accession Number 207223, the cDNA of ATCC[®] Accession Number PTA-248, or the cDNA of ATCC[®] Accession Number PTA-293, or a complement thereof.

In other embodiments, the isolated nucleic acid molecules encode an extracellular, transmembrane, or cytoplasmic domain of a polypeptide of the invention.

In another embodiment, the invention provides an isolated nucleic acid molecule which is antisense to the coding strand of a nucleic acid of the invention.

Another aspect of the invention provides vectors, *e.g.*, recombinant expression vectors, comprising a nucleic acid molecule of the invention. In another embodiment, the invention provides host cells containing such a vector or a nucleic acid molecule of the invention. The invention also provides methods for producing a polypeptide of the invention by culturing, in a suitable medium, a host cell of the invention containing a recombinant expression vector such that a polypeptide is produced.

Another aspect of this invention features isolated or recombinant proteins and polypeptides of the invention. Preferred proteins and polypeptides possess at least one biological activity possessed by the corresponding naturally-occurring

human polypeptide. An activity, a biological activity, or a functional activity of a polypeptide or nucleic acid of the invention refers to an activity exerted by a protein, polypeptide or nucleic acid molecule of the invention on a responsive cell as determined *in vivo*, or *in vitro*, according to standard techniques. Such activities can
5 be a direct activity, such as an association with or an enzymatic activity on a second protein or an indirect activity, such as a cellular signaling activity mediated by interaction of the protein with a second protein.

For TANGO 244, biological activities include, *e.g.*, (1) the ability to form protein-protein interactions with proteins in the signaling pathway of the naturally-
10 occurring polypeptide; (2) the ability to bind a ligand of the naturally-occurring polypeptide; and (3) the ability to interact with a TANGO 244 receptor. Other activities include the ability to modulate function, survival, morphology, proliferation and/or differentiation of cells of tissues in which it is expressed.

For TANGO 246, biological activities include, *e.g.*, (1) the ability to form
15 protein-protein interactions with proteins in the signaling pathway of the naturally-occurring polypeptide; (2) the ability to bind a ligand of the naturally-occurring polypeptide; and (3) the ability to interact with a TANGO 246 receptor. Other activities include the ability to modulate function, survival, morphology, proliferation and/or differentiation of cells of tissues in which it is expressed. TANGO 246
20 biological activities can include the ability to act as a small molecule transporter or a cell cycle regulator.

For TANGO 275, biological activities include, *e.g.*, (1) the ability to form protein-protein interactions with proteins in the signaling pathway of the naturally-occurring polypeptide; (2) the ability to bind a ligand of the naturally-occurring
25 polypeptide; and (3) the ability to interact with a TANGO 275 receptor. Other activities include the ability to modulate function, survival, morphology, proliferation and/or differentiation of cells of tissues in which it is expressed (*e.g.*, cells of the pituitary gland). TANGO 275 biological activities can include: (1) the ability to act as a TGF- β binding protein; (2) the ability to facilitate the normal assembly and
30 secretion of large latent complexes containing TGF- β ; (3) the ability to target latent TGF- β to connective tissue; (4) the ability to target latent TGF- β to the cell surface;

(5) the ability to modulate bone formation, renewal, or remodelling; and (6) the ability to modulate the development or function of the heart, cardiovascular system, brain, placenta, liver, skeletal muscle, kidney or pancreas.

For TANGO 300, biological activities include, *e.g.*, (1) the ability to form
5 protein-protein interactions with proteins in the signaling pathway of the naturally-
occurring polypeptide; (2) the ability to bind a ligand of the naturally-occurring
polypeptide; (3) the ability to interact with a TANGO 300 receptor; and (4) the ability
to mediate an intracellular signal. Other activities include the ability to modulate
function, survival, morphology, proliferation and/or differentiation of cells of tissues
10 in which it is expressed.

For MANGO 245, biological activities include, *e.g.*, (1) the ability to form
protein-protein interactions with proteins in the signaling pathway of the naturally-
occurring polypeptide; (2) the ability to bind a ligand of the naturally-occurring
polypeptide; and (3) the ability to interact with a MANGO 245 receptor. Other
15 activities include the ability to modulate function, survival, morphology, proliferation
and/or differentiation of cells of tissues in which it is expressed, *e.g.*, the central
nervous system, and the ability to modulate the cellular functions of cells of the
nervous system (neurons and glial cells), and the ability to act as a modulator of
complement function.

20 In one embodiment, a polypeptide of the invention has an amino acid
sequence sufficiently identical to an identified domain of a polypeptide of the
invention. As used herein, the term "sufficiently identical" refers to a first amino acid
or nucleotide sequence which contains a sufficient or minimum number of identical or
equivalent (*e.g.*, with a similar side chain) amino acid residues or nucleotides to a
25 second amino acid or nucleotide sequence such that the first and second amino acid or
nucleotide sequences have a common structural domain and/or common functional
activity. For example, amino acid or nucleotide sequences which contain a common
structural domain having about 60% identity, preferably 65% identity, more
preferably 75%, 85%, 95%, 98% or more identity are defined herein as sufficiently
30 identical.

In one embodiment, a TANGO 244, TANGO 246, TANGO 275, TANGO 300, or MANGO 245 polypeptide of the invention includes a signal peptide.

In another embodiment, a nucleic acid molecule of the invention encodes a TANGO 244, TANGO 246, TANGO 275, TANGO 300, or MANGO 245 polypeptide
5 which includes a signal peptide. In another embodiment, a TANGO 244, TANGO 246, or MANGO 245 polypeptide of the invention also includes one or more of the following domains: (1) an extracellular domain; (2) a transmembrane domain; and (3) a cytoplasmic domain.

In one embodiment, the isolated polypeptide of the invention lacks both a
10 transmembrane and a cytoplasmic domain. In another embodiment, the polypeptide lacks both a transmembrane domain and a cytoplasmic domain and is soluble under physiological conditions.

The polypeptides of the present invention, or biologically active portions thereof can be operably linked to a heterologous amino acid sequence to form fusion
15 proteins. In one embodiment, the fusion protein consists of a chimeric protein assembled from portions of the protein from different species. In another embodiment, the fusion protein consists of the amino terminal portion of murine MANGO 245 attached to the carboxy terminal portion of human MANGO 245.

The invention further features antibodies that specifically bind a
20 polypeptide of the invention such as monoclonal or polyclonal antibodies. In addition, the polypeptides of the invention or biologically active portions thereof, or antibodies of the invention, can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides methods for detecting the
25 presence of the activity or expression of a polypeptide of the invention in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of activity such that the presence of activity is detected in the biological sample.

In another aspect, the invention provides methods for modulating activity
30 of a polypeptide of the invention comprising contacting a cell with an agent that modulates (inhibits or stimulates) the activity or expression of a polypeptide of the

invention such that activity or expression in the cell is modulated. In one embodiment, the agent is an antibody that specifically binds to a polypeptide of the invention.

In another embodiment, the agent modulates expression of a polypeptide
5 of the invention by modulating transcription, splicing, or translation of an mRNA encoding a polypeptide of the invention. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of an mRNA encoding a polypeptide of the invention.

The present invention also provides methods to treat a subject having a
10 disorder characterized by aberrant activity of a polypeptide of the invention or aberrant expression of a nucleic acid of the invention by administering an agent which is a modulator of the activity of a polypeptide of the invention or a modulator of the expression of a nucleic acid of the invention to the subject. In one embodiment, the modulator is a protein of the invention. In another embodiment, the modulator is a
15 nucleic acid of the invention. In other embodiments, the modulator is a peptide, peptidomimetic, or other small organic molecule.

The present invention also provides diagnostic assays for identifying the presence or absence of a genetic lesion or mutation characterized by at least one of:
(i) aberrant modification or mutation of a gene encoding a polypeptide of the
20 invention, (ii) mis-regulation of a gene encoding a polypeptide of the invention, and (iii) aberrant post-translational modification of the invention wherein a wild-type form of the gene encodes a protein having the activity of the polypeptide of the invention.

In another aspect, the invention provides a method for identifying a compound that binds to or modulates the activity of a polypeptide of the invention. In
25 general, such methods entail measuring a biological activity of the polypeptide in the presence and absence of a test compound and identifying those compounds which alter the activity of the polypeptide.

The invention also features methods for identifying a compound which modulates the expression of a polypeptide or nucleic acid of the invention by
30 measuring the expression of the polypeptide or nucleic acid in the presence and absence of the compound.

In yet a further aspect, the invention provides substantially purified antibodies or fragments thereof including human and non-human antibodies or fragments thereof which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23 or 91, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207220, 207223, PTA-248 or PTA-293; a fragment of at least 15 amino acid residues of the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91; an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID NO:2, 5, 8, 11, 14, 17, 20, 23, or 91, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 or the BESTFIT program with BLOSUM 62 scoring matrix, gap open penalty of 12, frame shift penalty of 5, gap extend penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of SEQ ID:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. In various embodiments, the substantially purified antibodies of the invention, or fragments thereof can be human, non-human, chimeric and/or humanized antibodies.

Any of the antibodies of the invention can be conjugated to a therapeutic moiety or to a detectable substance. Non-limiting examples of detectable substances that can be conjugated to the antibodies of the invention are an enzyme, a prosthetic group, a fluorescent material, a luminescent material, a bioluminescent material, and a radioactive material.

The invention also provides a kit containing an antibody of the invention conjugated to a detectable substance, and instructions for use. Still another aspect of the invention is a pharmaceutical composition comprising an antibody of the invention and a pharmaceutically acceptable carrier. In preferred embodiments, the pharmaceutical composition contains an antibody of the invention, a therapeutic moiety, and a pharmaceutically acceptable carrier.

Other features and advantages of the invention will be apparent from the following detailed description and Claims.

Brief Description of the Drawings

5 *Figure 1* depicts the cDNA sequence (SEQ ID NO:1) and the predicted amino acid sequence (SEQ ID NO:2) of human TANGO 244. The open reading frame of SEQ ID NO:1 extends from nucleotide 85 to nucleotide 570 of SEQ ID NO:1 (SEQ ID NO:3).

Figure 2 depicts a hydropathy plot of human TANGO 244. Relatively
10 hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace.

Figure 3 depicts an alignment of the immunoglobulin domain of human TANGO 244 (SEQ ID NO:28) with a consensus hidden Markov model
15 immunoglobulin domain (SEQ ID NO:29). In the consensus sequence, more conserved residues are indicated by uppercase letters, and less conserved residues are indicated by lowercase letters. A “-” within a sequence indicates a gap created in the sequence for purposes of alignment. A “+” between the aligned sequences indicates a conservative amino acid difference.

20 *Figure 4* depicts an alignment of the amino acid sequence of human TANGO 244 (SEQ ID NO:2) and the amino acid sequence of human CTH (SEQ ID NO:81; Genbank Accession Number AF061022; Marcuz et al., *Eur J. Immunol.* 28:4094-4104). This alignment was created using ALIGN (version 2.0; PAM120 scoring matrix; gap length penalty of 12; gap penalty of 4). In this alignment, the
25 sequences are 48.6% identical.

Figures 5A-5B depict the cDNA sequence (SEQ ID NO:4) and the predicted amino acid sequence (SEQ ID NO:5) of human TANGO 246. The open reading frame of SEQ ID NO:4 extends from nucleotide 94 to nucleotide 1080 of SEQ ID NO:4 (SEQ ID NO:6).

30 *Figure 6* depicts a hydropathy plot of human TANGO 246. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic

regions are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace.

Figure 7 depicts an alignment of the cell cycle protein domain of human TANGO 246 (SEQ ID NO:30) with a consensus hidden Markov model cell cycle protein domain (SEQ ID NO:31). In the consensus sequence, more conserved residues are indicated by uppercase letters, and less conserved residues are indicated by lowercase letters. A “-” within a sequence indicates a gap created in the sequence for purposes of alignment. A “+” between the aligned sequences indicates a conservative amino acid difference.

Figure 8 depicts an alignment of the ABC transporter domain of human TANGO 246 (SEQ ID NO:32) with a consensus hidden Markov model ABC transporter domain (SEQ ID NO:33). In the consensus sequence, more conserved residues are indicated by uppercase letters, and less conserved residues are indicated by lowercase letters. A “-” within a sequence indicates a gap created in the sequence for purposes of alignment. A “+” between the aligned sequences indicates a conservative amino acid difference.

Figures 9A-9D depict the cDNA sequence (SEQ ID NO:7) and the predicted amino acid sequence (SEQ ID NO:8) of human TANGO 275. The open reading frame of SEQ ID NO:7 extends from nucleotide 23 to nucleotide 3931 SEQ ID NO:7 (SEQ ID NO:9).

Figure 10 depicts a hydropathy plot of human TANGO 275. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace.

Figures 11A-11B depict alignments of the EGF-like domains of human TANGO 275 (SEQ ID NOs:34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47 and 48) with a consensus hidden Markov model EGF-like domain (SEQ ID NO:49). In the consensus sequence, more conserved residues are indicated by uppercase letters, and less conserved residues are indicated by lowercase letters. A “-” within a

sequence indicates a gap created in the sequence for purposes of alignment. A "+" between the aligned sequences indicates a conservative amino acid difference.

Figure 12 depicts alignments of the TB domains of human TANGO 275 (SEQ ID NOs:50, 51, 52, and 53) with a consensus hidden Markov model TB domain (SEQ ID NO:54). In the consensus sequence, more conserved residues are indicated by uppercase letters, and less conserved residues are indicated by lowercase letters. A "-" within a sequence indicates a gap created in the sequence for purposes of alignment. A "+" between the aligned sequences indicates a conservative amino acid difference.

Figure 13 depicts alignments of the metallothionein domain of human TANGO 275 (SEQ ID NO:55) with a consensus hidden Markov model metallothionein domain (SEQ ID NO:56). In the consensus sequence, more conserved residues are indicated by uppercase letters, and less conserved residues are indicated by lowercase letters. A "-" within a sequence indicates a gap created in the sequence for purposes of alignment. A "+" between the aligned sequences indicates a conservative amino acid difference.

Figures 14A-14H depict an alignment of the nucleotide sequence of human TANGO 275 (SEQ ID NO:7) and the nucleotide sequence of murine LTBP-3 (Genbank Accession Number L40459; SEQ ID NO:82). This alignment was created using ALIGN (version 2.0; PAM120 scoring matrix; gap length penalty of 12; gap penalty of 4). In this alignment, the sequences are 77.1% identical.

Figures 15A-15C depict an alignment of the amino acid sequence of human TANGO 275 (SEQ ID NO:8) and the amino acid sequence of murine LTBP-3 (GENSEQ Accession Number R79475; SEQ ID NO:83). This alignment was created using ALIGN (version 2.0; PAM120 scoring matrix, gap length penalty of 12; gap penalty of 4). In this alignment, the sequences are 82.8% identical.

Figures 16A-16G depict the cDNA sequence (SEQ ID NO:10) and the predicted amino acid sequence (SEQ ID NO:11) of murine TANGO 275. The open reading frame of SEQ ID NO:10 extends from nucleotide 157 to nucleotide 3916 of SEQ ID NO:10 (SEQ ID NO:12).

Figure 17A-17B depicts the cDNA sequence (SEQ ID NO:13) and the predicted amino acid sequence (SEQ ID NO:14) of human TANGO 300. The open reading frame of SEQ ID NO:13 extends from nucleotide 31 to nucleotide 1113 of SEQ ID NO:13 (SEQ ID NO:15).

5 *Figure 18* depicts a hydropathy plot of human TANGO 300. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace.

10 *Figures 19A-19C* depicts the cDNA sequence (SEQ ID NO:16) and the predicted amino acid sequence (SEQ ID NO:17) of murine TANGO 300. The open reading frame of SEQ ID NO:16 extends from nucleotide 41 to nucleotide 1195 of SEQ ID NO:16 (SEQ ID NO:18).

15 *Figure 20* depicts a hydropathy plot of murine TANGO 300. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace.

20 *Figures 21A-21B* depict an alignment of the nucleotide sequence of the ORF of human TANGO 300 (SEQ ID NO:15) and the nucleotide sequence of the ORF of murine TANGO 300 (SEQ ID NO:18). This alignment was created using BESTFIT (BLOSUM 62 scoring matrix; gap open penalty of 12; frame shift penalty of 5; gap extend penalty of 4). In this alignment, the sequences are 77.7% identical.

25 *Figure 22* depicts an alignment of the amino acid sequence of human TANGO 300 (SEQ ID NO:14) and the amino acid sequence of murine TANGO 300 (SEQ ID NO:17). This alignment was created using BESTFIT (BLOSUM 62 scoring matrix; gap open penalty of 12; frame shift penalty of 5; gap extend penalty of 4). In this alignment, the sequences are 69.6% identical.

30 *Figures 23A-23B* depicts the cDNA sequence (SEQ ID NO:19) and the predicted amino acid sequence (SEQ ID NO:20) of human MANGO 245. The open

reading frame of SEQ ID NO:19 extends from nucleotide 105 to nucleotide 1148 of SEQ ID NO:19 (SEQ ID NO:21).

Figure 24 depicts a hydropathy plot of human MANGO 245. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace.

Figures 25A-25B depict the cDNA sequence (SEQ ID NO:22) and the predicted amino acid sequence (SEQ ID NO:23) of monkey MANGO 245. The open reading frame of SEQ ID NO:22 extends from nucleotide 250 to nucleotide 1236 of SEQ ID NO:22 (SEQ ID NO:24).

Figure 26 depicts an alignment of the amino acid sequences of human MANGO 245 (SEQ ID NO:20) and monkey MANGO 245 (SEQ ID NO:23). This alignment was created using ALIGN (version 2.0; PAM120 scoring matrix, gap length penalty of 12; gap penalty of 4). In this alignment, the sequences are 84.8% identical.

Figure 27 depicts alignments of the CIq domains of human MANGO 245 (SEQ ID NOs:70 and 71) with a consensus hidden Markov model CIq domain (SEQ ID NO:72). In the consensus sequence, more conserved residues are indicated by uppercase letters, and less conserved residues are indicated by lowercase letters. A "-" within a sequence indicates a gap created in the sequence for purposes of alignment. A "+" between the aligned sequences indicates a conservative amino acid difference.

Figure 28 depicts alignments of the CIq domains of monkey MANGO 245 (SEQ ID NOs:73 and 74) with a consensus hidden Markov model CIq domain (SEQ ID NO:72). In the consensus sequence, more conserved residues are indicated by uppercase letters, and less conserved residues are indicated by lowercase letters. A "-" within a sequence indicates a gap created in the sequence for purposes of alignment. A "+" between the aligned sequences indicates a conservative amino acid difference.

Figure 29 depicts the cDNA sequence (SEQ ID NO:25) of murine MANGO 245 and the predicted amino acid sequence (SEQ ID NO:91) of murine

MANGO 245. The open reading frame of SEQ ID NO:25 extends from nucleotide 29 to nucleotide 625 of SEQ ID NO:25 (SEQ ID NO:92).

Figures 30A-30B depict an alignment of nucleotide 51 to nucleotide 748 of human MANGO 245 (SEQ ID NO:19) with murine MANGO 245 (SEQ ID NO:25).

- 5 This alignment was created using BESTFIT (BLOSUM 62 scoring matrix; gap open penalty of 12; frame shift penalty of 5; gap extend penalty of 4). In this alignment, the sequences are 89.6% identical.

Figure 31 depicts an alignment of the amino acid sequence of human TANGO 246 (SEQ ID NO:5) and the amino acid sequence of *Arabidopsis thaliana*
10 AIG1 (Genbank Accession Number AAC49289; SEQ ID NO:87).

Figure 32A-32B depicts an alignment of the amino acid sequence of murine TANGO 275 (SEQ ID NO:11) and the amino acid sequence of murine LTBP-3 (GENSEQ® Accession Number R79475; SEQ ID NO:83). This alignment was created using ALIGN (version 2.0; PAM120 scoring matrix, gap length penalty of 12;
15 gap penalty of 4). In this alignment, the sequences are 97.4% identical.

Detailed Description of the Invention

The present invention is based, at least in part, on the discovery of cDNA molecules encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300, and
20 MANGO 245, all of which are predicted to be either wholly secreted or transmembrane proteins.

The proteins and nucleic acid molecules of the present invention comprise a family of molecules having certain conserved structural and functional features. As used herein, the term “family” is intended to mean two or more proteins or nucleic
25 acid molecules having a common structural domain and having sufficient amino acid or nucleotide sequence identity as defined herein. Family members can be from either the same or different species. For example, a family can comprise two or more proteins of human origin, or can comprise one or more proteins of human origin and one or more of non-human origin. For example, the human MANGO 245 and
30 monkey MANGO 245 genes described herein are both members of the MANGO 245

family. Two different polypeptides encoded by splice variants of a given transcript are also considered members of the same family.

Members of the same family may also have common structural domains.

For example, a TANGO 244, TANGO 246, TANGO 275, TANGO 300, or MANGO
5 245 family member includes a signal peptide. As used herein, a "signal peptide" includes a peptide of at least about 15 amino acid residues in length which occurs at the N-terminus of secretory and membrane-bound proteins and which contains at least about 70% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. The sequence can contain
10 about 15 to 45 amino acid residues or about 17 to 22 amino acid residues, and has at least about 60-80%, 65-75%, or about 70% hydrophobic residues. A signal peptide serves to direct a protein containing such a sequence to a lipid bilayer.

Thus, in one embodiment, a TANGO 244 protein contains a signal peptide of about amino acids 1 to 26 (1 to 24, 1 to 25, 1 to 27, or 1 to 28) of SEQ ID NO:2
15 (SEQ ID NO:26). In one embodiment, a TANGO 275 protein contains a signal peptide of about amino acids 1 to 29 (1 to 27, 1 to 28, 1 to 30, 1 to 31) of SEQ ID NO:8 (SEQ ID NO:60). In one embodiment, a TANGO 300 protein contains a signal peptide of about amino acids 1 to 19 (1 to 17, 1 to 18, 1 to 20, 1 to 21) of SEQ ID NO:14 or SEQ ID NO:17 (SEQ ID NO:62 and SEQ ID NO:64, respectively). In one
20 embodiment, a MANGO 245 protein contains a signal peptide of amino acids 1 to 16 (1 to 14, 1 to 15, 1 to 17, 1 to 18) of SEQ ID NO:20 or SEQ ID NO:23 (SEQ ID NO:66 and SEQ ID NO:68, respectively).

The signal peptide is cleaved during processing of the mature protein. Sometimes the initial methionine residue is also cleaved from the protein during
25 signal peptide processing. Thus, in one embodiment, a TANGO 244 protein does not contain a signal peptide or an initial methionine residue and begins from residue 2 of SEQ ID NO:2. In one embodiment, a TANGO 275 protein does not contain a signal peptide or an initial methionine residue and begins from residue 2 of SEQ ID NO:8. In one embodiment, a TANGO 300 protein does not contain a signal peptide or an
30 initial methionine residue and begins from residue 2 of SEQ ID NO:14 or SEQ ID NO:17. Thus, in one embodiment, a MANGO 245 protein does not contain a signal

peptide or an initial methionine residue and begins from residue 2 of SEQ ID NO:20 or SEQ ID NO:23.

In some embodiments of the invention, the domains and the mature protein resulting from the cleavage of such signal peptides are also included herein. For example, the cleavage of a signal peptide consisting of amino acids 1 to 26 of SEQ ID NO:2 (SEQ ID NO:26) results in a mature TANGO 244 protein corresponding to amino acids 27-162 of SEQ ID NO:2 (SEQ ID NO:27). The signal peptide is normally cleaved during processing of the mature protein.

In another example, a TANGO 244, TANGO 246 or MANGO 245 family member also includes one or more of the following domains: (1) an extracellular domain; (2) a transmembrane domain; and (3) a cytoplasmic domain as described herein.

TANGO 244 family members can also include an immunoglobulin domain. Immunoglobulin domains are present in a variety of proteins and are involved in protein-protein and protein-ligand interaction at the cell surface. A consensus hidden Markov model immunoglobulin domain has the sequence of SEQ ID NO:29. This consensus sequence is shown in Figure 3 where the more conserved residues in the consensus sequence are indicated by uppercase letters and the less conserved residues in the consensus sequence are indicated by lowercase letters. Human TANGO 244 includes an immunoglobulin domain at amino acids 37 to 97 of SEQ ID NO:2 (SEQ ID NO:28).

TANGO 246 family members can also include a cell cycle protein domain. A consensus hidden Markov model cell cycle protein domain has the sequence of SEQ ID NO:31. This consensus sequence is shown in Figure 7 where the more conserved residues in the consensus sequence are indicated by uppercase letters and the less conserved residues in the consensus sequence are indicated by lowercase letters. Human TANGO 246 includes a cell cycle protein domain at amino acids 27 to 215 of SEQ ID NO:5 (SEQ ID NO:30). Among the proteins which have a cell cycle protein domain are CDC3, CDC10, and CDC11, all of which are important for regulation of the cell cycle. Many proteins which include this domain are GTP binding proteins.

In addition, TANGO 246 family members can also include an ABC transporter domain. A consensus hidden Markov model ABC transporter protein domain has the sequence of SEQ ID NO:33. This consensus sequence is shown in Figure 8 where the more conserved residues in the consensus sequence are indicated by uppercase letters and the less conserved residues in the consensus sequence are indicated by lowercase letters. The ABC transporter protein domain of TANGO 246 is located at amino acids 30 to 192 of SEQ ID NO:5 (SEQ ID NO:32). A number of proteins having an ABC transporter protein domain act as active transporters of small hydrophilic molecules (*e.g.*, ions) across cell membranes, including intracellular membranes. In eukaryotes, ABC transporter protein domains are present in multidrug resistance proteins. These protein are involved in extrusion of drugs from cells and play a key role in drug resistance. This domain is also present in cystic fibrosis transmembrane conductance regulator (CFTR), a protein that likely acts as a chloride ion transporter. Many proteins having an ABC transporter domain are ATP binding proteins.

TANGO 275 family members can include an EGF-like domain. A consensus hidden Markov model EGF-like domain has the sequence of SEQ ID NO:49. This consensus sequence is shown in Figures 11A-11B where the more conserved residues in the consensus sequence are indicated by uppercase letters and the less conserved residues in the consensus sequence are indicated by lowercase letters. Human TANGO 275 includes EFG-like domains at amino acids 99 to 126 (SEQ ID NO:34), 345 to 380 (SEQ ID NO:35), 564 to 600 (SEQ ID NO:36), 606 to 644 (SEQ ID NO:37), 650 to 687 (SEQ ID NO:38), 693 to 728 (SEQ ID NO:39), 734 to 769 (SEQ ID NO:40), 775 to 810 (SEQ ID NO:41), 816 to 850 (SEQ ID NO:42), 856 to 893 (SEQ ID NO:43), 983 to 1020 (SEQ ID NO:44), 1026 to 1061 (SEQ ID NO:45), 1072 to 1107 (SEQ ID NO:46), 1203 to 1238 (SEQ ID NO:47), and 1244 to 1283 (SEQ ID NO:48). One or more EGF-like domains (*e.g.*, 1, 2, 4, 8, 13, 17, or 44 copies) are found in the extracellular domain of a wide range of proteins of transmembrane and wholly secreted proteins having diverse function. The consensus EGF-like domain sequence includes six cysteines, all of which are thought to be involved in disulfide bonds.

TANGO 275 family members can include a transforming growth factor β binding protein-like domains (TB domains). A consensus hidden Markov model TB domain has the amino acid sequence of SEQ ID NO:54. This consensus sequence is shown in Figure 12 where the more conserved residues in the consensus sequence are indicated by uppercase letters and the less conserved residues in the consensus sequence are indicated by lowercase letters. Human TANGO 275 includes TB domains at amino acids 273 to 316 (SEQ ID NO:50), 399 to 440 (SEQ ID NO:51), 913 to 956 (SEQ ID NO:52), and 1132 to 1177 (SEQ ID NO:53) of SEQ ID NO:8. A TB domain is found in matrix fibrils (Yuan et al., 1997, *EMBO J.* 16:6659-66).

10 TANGO 275 family members can include a metallothionein domain. A consensus hidden Markov model metallothionein domain has the amino acid sequence of SEQ ID NO:56. This consensus sequence is shown in Figure 13 where the more conserved residues in the consensus sequence are indicated by uppercase letters and the less conserved residues in the consensus sequence are indicated by lowercase letters. Human TANGO 275 includes a metallothionein domain at amino acids 794 to 708 (SEQ ID NO:55) of SEQ ID NO:8. Metallothionein domains are found in proteins which bind heavy metals (*e.g.*, copper, zinc, cadmium, and nickel) through thiolate bonds.

MANGO 245 family members can also include a CIq domain. A consensus hidden Markov model CIq domain has the amino acid sequence of SEQ ID NO:72. This consensus sequence is shown in Figure 27 where the more conserved residues in the consensus sequence are indicated by uppercase letters and the less conserved residues in the consensus sequence are indicated by lowercase letters. Human MANGO 245 includes CIq domains at amino acids 31 to 156 of SEQ ID NO:20 (SEQ ID NO:70) and amino acids 178 to 294 of SEQ ID NO:20 (SEQ ID NO:71). Monkey MANGO 245 includes CIq domains at amino acids 31 to 156 of SEQ ID NO:23 (SEQ ID NO:73) and amino acids 178 to 311 of SEQ ID NO: (SEQ ID NO:74). Murine MANGO 245 includes a CIq domain at amino acids 30 to 155 of SEQ ID NO: 91 (SEQ ID NO:93). CIq domains are found in wholly secreted or membrane bound proteins that are short-chain collagens and collagen-like molecules. The domain likely forms ten β -strands interspersed by β -turns and/or loops.

Various features of TANGO 244, TANGO 246, TANGO 275, TANGO 300, and MANGO 245 are summarized below.

TANGO 244

5 A cDNA encoding TANGO 244 was identified by analyzing the sequences of clones present in a human fetal lung cDNA library.

This analysis led to the identification of a clone, Athua62f9, encoding full-length human TANGO 244. The cDNA of this clone is 1513 nucleotides long (Figure 1; SEQ ID NO:1). The 486 nucleotide open reading frame of this cDNA, nucleotide
10 85 to nucleotide 570 of SEQ ID NO:1 (SEQ ID NO:3), encodes a 162 amino acid protein (Figure 1; SEQ ID NO:2).

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that human TANGO 244 includes a 26 amino acid signal peptide (amino acid 1 to about amino acid 26 of SEQ ID NO:2; SEQ ID
15 NO:26) preceding the mature human TANGO 244 protein (corresponding to about amino acid 27 to amino acid 162 of SEQ ID NO:2; SEQ ID NO:27).

Human TANGO 244 is a transmembrane protein having an extracellular domain which extends from about amino acid 27 to about amino acid 119 of SEQ ID NO:2 (SEQ ID NO:75), a transmembrane domain which extends from about amino
20 acid 120 to about amino acid 142 of SEQ ID NO:2 (SEQ ID NO:76), and a cytoplasmic domain which extends from about amino acid 143 to amino acid 162 of SEQ ID NO:2 (SEQ ID NO:77).

Alternatively, in another embodiment, a human TANGO 244 protein contains an extracellular domain at amino acid residues 143 to 162 of SEQ ID NO:2
25 (SEQ ID NO:77), transmembrane domains at amino acid residues 120 to 142 of SEQ ID NO:2 (SEQ ID NO:76), and a cytoplasmic domain at amino acid residues 27 to 119 of SEQ ID NO:2 (SEQ ID NO:75).

Human TANGO 244 that has not been post-translationally modified is predicted to have a molecular weight of 16.8 kDa prior to cleavage of its signal
30 peptide and a molecular weight of 14.2 kDa subsequent to cleavage of its signal peptide.

Human TANGO 244 includes an immunoglobulin domain at amino acids 37 to 97 of SEQ ID NO:2 (SEQ ID NO:28). Figure 3 depicts an alignment of the immunoglobulin domain of human TANGO 244 with a consensus hidden Markov model immunoglobulin domain derived from a (SEQ ID NO:29).

5 Within human TANGO 244, an N-glycosylation site is present at amino acids 84 to 87 of SEQ ID NO:2. A protein kinase C phosphorylation sites is present at amino acids 92 to 94 of SEQ ID NO:2. N-myristylation sites are present at amino acids 11 to 16, 37 to 42, 91 to 96, 102 to 107, and 122 to 127 of SEQ ID NO:2. An amidation site is present at amino acids 148 to 151 of SEQ ID NO:2.

10 Clone Athua62f9, which encodes human TANGO 244, was deposited as EpT244 with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on April 21, 1999 and assigned Accession Number 207223. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the
15 Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 2 depicts a hydropathy plot of human TANGO 244. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions
20 are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 2 indicates that human TANGO 244 has a signal peptide at its amino terminus and an internal hydrophobic region, suggesting that TANGO 244 is a transmembrane protein.

Northern blot analysis of human TANGO 244 expression revealed that
25 human TANGO 244 is expressed in the colon, kidney, liver, and lung.

Human TANGO 244 has sequence homology to human CTH (Marcuz et al., 1998, *Eur. J. Immunol.* 28:4094-4104; Genbank Accession Number AFO61022). Figure 4 depicts an alignment of the amino acid sequence of human TANGO 244 (SEQ ID NO:2) and the amino acid sequence of human CTH (SEQ ID NO:81). In
30 this alignment, the sequences are 48.6% identical overall. However, there is a

substantial region of complete identity. TANGO 244 may act as a immunoglobulin superfamily-type receptor.

Use of TANGO 244 Nucleic Acids, Polypeptides, and Modulators Thereof

5 TANGO 244 polypeptides, nucleic acids, and modulators thereof can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues in which they are expressed. Such molecules can be used to treat disorders associated with abnormal or aberrant metabolism or function of cells in the tissues in which they are expressed. Tissues in which TANGO 244 is expressed
10 include, for example, the colon, kidney, liver, and lung. Such disorders include but are limited to lymphoma, leukemia, amyloidosis, scleroderma, mastocytosis.

In one example, TANGO 244 polypeptides, nucleic acids, or modulators thereof can be used to treat colonic disorders, such as congenital anomalies (*e.g.*, megacolon and imperforate anus), idiopathic disorders (*e.g.*, diverticular disease and
15 melanosis coli), vascular lesions (*e.g.*, ischemic colitis, hemorrhoids, angiodysplasia), inflammatory diseases (*e.g.*, idiopathic ulcerative colitis, pseudomembranous colitis, and lymphopathia venereum), tumors (*e.g.*, hyperplastic polyps, adenomatous polyps, bronchogenic cancer, colonic carcinoma, squamous cell carcinoma, adenoacanthomas, sarcomas, lymphomas, argentaffinomas, carcinoids,
20 and melanocarcinomas) and Crohn's Disease.

In another example, TANGO 244 polypeptides, nucleic acids, or modulators thereof can be used to treat renal disorders, such as glomerular diseases (*e.g.*, acute and chronic glomerulonephritis, rapidly progressive glomerulonephritis, nephrotic syndrome, focal proliferative glomerulonephritis, glomerular lesions
25 associated with systemic disease, such as systemic lupus erythematosus, Goodpasture's syndrome, multiple myeloma, diabetes, neoplasia, sickle cell disease, and chronic inflammatory diseases), tubular diseases (*e.g.*, acute tubular necrosis and acute renal failure, polycystic renal disease, medullary sponge kidney, medullary cystic disease, nephrogenic diabetes, and renal tubular acidosis), tubulointerstitial
30 diseases (*e.g.*, pyelonephritis, drug and toxin induced tubulointerstitial nephritis, hypercalcemic nephropathy, and hypokalemic nephropathy) acute and rapidly

progressive renal failure, chronic renal failure, nephrolithiasis, gout, vascular diseases (e.g., hypertension and nephrosclerosis, microangiopathic hemolytic anemia, atheroembolic renal disease, diffuse cortical necrosis, and renal infarcts), or tumors (e.g., renal cell carcinoma and nephroblastoma).

- 5 In another example, TANGO 244 polypeptides, nucleic acids, or modulators thereof can be used to treat hepatic (liver) disorders, such as jaundice, hepatic failure, hereditary hyperbilirubinemias (e.g., Gilbert's syndrome, Crigler-Naijar syndromes and Dubin-Johnson and Rotor's syndromes), hepatic circulatory disorders (e.g., hepatic vein thrombosis and portal vein obstruction and thrombosis) hepatitis
- 10 (e.g., chronic active hepatitis, acute viral hepatitis, and toxic and drug-induced hepatitis) cirrhosis (e.g., alcoholic cirrhosis, biliary cirrhosis, and hemochromatosis), or malignant tumors (e.g., primary carcinoma, hepatoma, hepatoblastoma, liver cysts and angiosarcoma).

- In another example, TANGO 244 polypeptides, nucleic acids, or
- 15 modulators thereof can be used to treat pulmonary (lung) disorders, such as atelectasis, cystic fibrosis, rheumatoid lung disease, pulmonary congestion or edema, chronic obstructive airway disease (e.g., emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis), diffuse interstitial diseases (e.g., sarcoidosis, pneumoconiosis, hypersensitivity pneumonitis, bronchiolitis Goodpasture's
- 20 syndrome, idiopathic pulmonary hemosiderosis, idiopathic pulmonary fibrosis, pulmonary alveolar proteinosis, desquamative interstitial pneumonitis, chronic interstitial pneumonia, fibrosing alveolitis, hamman-rich syndrome, pulmonary eosinophilia, diffuse interstitial fibrosis, Wegener's granulomatosis, lymphomatoid granulomatosis, and lipid pneumonia), or tumors (e.g., bronchogenic carcinoma,
- 25 bronchioloalveolar carcinoma, bronchial carcinoid, hamartoma, and mesenchymal tumors).

- Because TANGO 244 includes immunoglobulin domains and has homology to human CTH, TANGO 244 polypeptides, nucleic acids, and modulators thereof can be used to treat disorders involving an immune, allergic or autoimmune
- 30 response (e.g., arthritis, multiple sclerosis, meningitis, encephalitis, atherosclerosis, or infection).

Further, in light of TANGO 244's pattern of expression in humans, TANGO 244 expression can be utilized as a marker for specific tissues (*e.g.*, tissues of the colon, kidney, liver, or lung) and/or cells (*e.g.*, colon, renal, hepatic, or pulmonary) in which TANGO 244 is expressed. TANGO 244 nucleic acids can also
5 be utilized for chromosomal mapping.

TANGO 246

A cDNA encoding human TANGO 246 was identified by analyzing the sequences of clones present in a human fetal spleen cDNA library.

10 This analysis led to the identification of a clone, Athsa34d2, encoding full-length human TANGO 246. The cDNA of this clone is 1992 nucleotides long (Figures 5A-5B; SEQ ID NO:4). The 987 nucleotide open reading frame of this cDNA, nucleotide 94 to nucleotide 1080 of SEQ ID NO:4 (SEQ ID NO:6), encodes a 329 amino acid protein (Figures 5A-5B; SEQ ID NO:5).

15 Human TANGO 246 has a hydrophobic domain which extends from about amino acid 306 to about amino acid 323 of SEQ ID NO:5 (SEQ ID NO:58). This could represent a transmembrane domain or an internal signal peptide. This domain follows a domain which extends from about amino acid 1 to about amino acid 305 of SEQ ID NO:5 (SEQ ID NO:57) and is followed by a domain which extends from
20 about amino acid 324 to amino acid 329 of SEQ ID NO:5 (SEQ ID NO:59).

Human TANGO 246 includes a cell cycle protein domain at amino acids 27 to 215 of SEQ ID NO:5 (SEQ ID NO:30). Figure 7 depicts an alignment of the cell cycle protein domain of human TANGO 246 with a consensus hidden Markov model cell cycle protein domain (SEQ ID NO:31).

25 Human TANGO 246 includes an ABC transporter domain at amino acids 30 to 192 of SEQ ID NO:5 (SEQ ID NO:32). Figure 8 depicts an alignment of the ABC transporter domain of human TANGO 246 with a consensus hidden Markov model ABC transporter domain (SEQ ID NO:33).

Human TANGO 246 that has not been post-translationally modified is
30 predicted to have a molecular weight of 37.5 kDa.

Within human TANGO 246, a cAMP and cGMP-dependent protein kinase phosphorylation site is present at amino acids 71 to 74 of SEQ ID NO:5. Protein kinase C phosphorylation sites are present at amino acids 66 to 68, 75 to 77, 99 to 101, 134 to 136, 154 to 156, and 222 to 224 of SEQ ID NO:5. Casein kinase II phosphorylation sites are present at amino acids 75 to 78, 99 to 102, 127 to 130, 154 to 157, 194 to 197, and 299 to 302 of SEQ ID NO:5. A tyrosine kinase phosphorylation site is present at amino acids 214 to 221 of SEQ ID NO:5. N-myristylation sites are present at amino acids 40 to 45, 88 to 93, and 219 to 224 of SEQ ID NO:5. An ATP/GTP-binding site motif A is present at amino acids 37 to 44 of SEQ ID NO:5. An amidation site is present at amino acids 51 to 54 of SEQ ID NO:5.

Clone Athsa34d2, which encodes human TANGO 246, was deposited as EpT246 with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on April 21, 1999 and assigned Accession Number 207223. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 6 depicts a hydropathy plot of human TANGO 246. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 5 indicates the presence of a hydrophobic domain within human TANGO 246, suggesting that human TANGO 246 is either a transmembrane protein or a secreted protein which employs an internal signal peptide.

Human TANGO 246 has homology to *Arabidopsis thaliana* AIG1, a gene which is involved in resistance response (Genbank Accession Number AAC49289; Reuber and Ausubel, 1996, *Plant Cell* 8:241-249), and *Nicotiana tabacum* NTGP4 (Genbank Accession Number AAD09518). Figure 31 depicts an alignment of the amino acid sequence of human TANGO 246 (SEQ ID NO:5) and the amino acid

sequence of *Arabidopsis thaliana* AIG1 (Genbank Accession Number AAC49289 (SEQ ID NO:87)). In this alignment, the proteins are 31.2% identical.

Use of TANGO 246 Nucleic Acids, Polypeptides, and Modulators Thereof

5 TANGO 246 polypeptides, nucleic acids, and modulators thereof can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues in which they are expressed.

TANGO 246 includes an ABC transporter domain. Proteins having such a domain are involved in disorders of transport of small molecules across cell
10 membranes. Proteins having an ABC transporter domain are known to be involved in cystic fibrosis, hyperinsulinemia, adrenoleukodystrophy, familial intrahepatic cholestasis, sideroblastic anemia and ataxia, Stargardt disease, multidrug resistance, and hyperbilirubinemia II/Dubin-Johnson syndrome. Thus, TANGO 246 polypeptides, nucleic acids, and modulators thereof can be used to treat these and
15 other disorders.

TANGO 246 includes a cell cycle protein domain. Proteins having such a domain are involved in regulation of the cell cycle. Thus, TANGO 246 polypeptides, nucleic acids, and modulators thereof can be used to treat disorders such as Alzheimer's disease, vascular restinosis, polycystic kidney disease, transplant
20 rejection, chronic liver disease, and cancer.

Further, in light of TANGO 246's presence in a human fetal spleen cDNA library, TANGO 246 expression can be utilized as a marker for specific tissues (*e.g.*, lymphoid tissues such as the thymus and spleen) and/or cells (*e.g.*, lymphocytes and splenic) in which TANGO 246 is expressed. TANGO 246 nucleic acids can also be
25 utilized for chromosomal mapping.

TANGO 275

A cDNA encoding human TANGO 275 was identified by analyzing the sequences of clones present in a human pituitary gland cDNA library.

30 This analysis led to the identification of a clone, Athbb19d1, encoding full-length human TANGO 275. The cDNA of this clone is 4225 nucleotides long

(Figures 9A-9D; SEQ ID NO:7). The 3867 nucleotide open reading frame of this cDNA, nucleotide 65 to nucleotide 3931 of SEQ ID NO:7 (SEQ ID NO:9), encodes a 1289 amino acid protein (Figures 9A-9D; SEQ ID NO:8).

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that human TANGO 275 includes a 29 amino acid signal peptide (amino acid 1 to about amino acid 29 of SEQ ID NO:8; SEQ ID NO:60) preceding the mature human TANGO 275 protein (corresponding to about amino acid 30 to amino acid 1289 of SEQ ID NO:8; SEQ ID NO:61).

Human TANGO 275 that has not been post-translationally modified is predicted to have a molecular weight of 137.9 kDa prior to cleavage of its signal peptide and a molecular weight of 135.3 kDa subsequent to cleavage of its signal peptide.

Human TANGO 275 includes EFG-like domains at amino acids 99 to 126 (SEQ ID NO:34), 345 to 380 (SEQ ID NO:35), 564 to 600 (SEQ ID NO:36), 606 to 644 (SEQ ID NO:37), 650 to 687 (SEQ ID NO:38), 693 to 728 (SEQ ID NO:39), 734 to 769 (SEQ ID NO:40), 775 to 810 (SEQ ID NO:41), 816 to 850 (SEQ ID NO:42), 856 to 893 (SEQ ID NO:43), 983 to 1020 (SEQ ID NO:44), 1026 to 1061 (SEQ ID NO:45), 1072 to 1107 (SEQ ID NO:46), 1203 to 1238 (SEQ ID NO:47), and 1244 to 1283 (SEQ ID NO:48). An alignment of each of the EGF-like domains of human TANGO 275 with a consensus hidden Markov model EGF-like domain (SEQ ID NO:49) is shown in Figures 11A-11B.

Human TANGO 275 includes transforming growth factor β binding protein like domains (TB domains) at amino acids 273 to 316 (SEQ ID NO:50), 399 to 440 (SEQ ID NO:51), 913 to 956 (SEQ ID NO:52), and 1132 to 1177 (SEQ ID NO:53) of SEQ ID NO:8. An alignment of each of the TB domains of human TANGO 275 with a consensus hidden Markov model TB domain (SEQ ID NO:54) is shown in Figure 12.

Human TANGO 275 includes a metallothionein domain at amino acids 694 to 708 (SEQ ID NO:55) of SEQ ID NO:8. An alignment of the metallothionein domain of human TANGO 275 with a consensus hidden Markov model metallothionein domain (SEQ ID NO:56) is shown in Figure 13.

N-glycosylation sites are present at amino acids 75 to 78, 335 to 338, 831 to 834, 922 to 925, and 1261 to 1264 of SEQ ID NO:8.

Clone Athbb19d1, which encodes human TANGO 275, was deposited as EpT275 with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on April 21, 1999 and assigned Accession Number 207220. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 10 depicts a hydropathy plot of human TANGO 275. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 10 indicates that human TANGO 275 has a signal peptide at its amino terminus, suggesting that human TANGO 275 is a secreted protein.

Transcript analysis suggests that there are several splice variants of human TANGO 275.

Human TANGO 275 appears to be the human homolog of a murine latent transforming growth factor- β binding protein 3 (LTBP-3; Yin et al., J. Biol. Chem. 270:10147-60, 1995; Genbank Accession Number RL40459; PCT Application WO 95/22611; GENSEQ® Accession Number R79475). Figures 14A-14H depict an alignment of the nucleotide sequence of human TANGO 275 (SEQ ID NO:7) and the nucleotide sequence of murine LTBP-3 (Genbank Accession Number L40459; SEQ ID NO:82). This alignment was created using ALIGN (version 2.0; PAM120 scoring matrix; gap length penalty of 12; gap penalty of 4). In this alignment, the sequences are 77.1% identical. Figures 15A-15C depict an alignment of the amino acid sequence of human TANGO 275 (SEQ ID NO:8) and the amino acid sequence of murine LTBP-3 (GENSEQ® R79475; SEQ ID NO:83). This alignment was created using ALIGN (version 2.0; PAM120 scoring matrix; gap length penalty of 12; gap penalty of 4). In this alignment, the sequences are 82.8% identical.

Northern blot analysis of human TANGO 275 expression revealed that human TANGO 275 is expressed at a high level in the heart and at a moderate level in the brain, placenta, lung, liver, skeletal muscle, kidney and pancreas.

A murine TANGO 275 cDNA was identified. The cDNA of this clone is 5 4376 nucleotides long (Figures 16A-16G; SEQ ID NO:10). The 3759 nucleotide open reading frame of this cDNA, nucleotides of SEQ ID NO:10 (SEQ ID NO:12), encodes a 1253 amino acid protein (Figures 16A-16G; SEQ ID NO:11). Figures 32A-32B depict an alignment of the amino acid sequence encoded by this murine TANGO 275 cDNA clones (SEQ ID NO:11) and the amino acid sequence of murine LTBP-3 10 (GENSEQ® Accession Number R79475; SEQ ID NO:83). This alignment was created using ALIGN (version 2.0; PAM120 scoring matrix, gap length penalty of 12; gap penalty of 4). In this alignment, the sequences are 97.4% identical.

Use of TANGO 275 Nucleic Acids, Polypeptides, and Modulators Thereof

15 TANGO 275 polypeptides, nucleic acids, and modulators thereof can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues in which they are expressed. Such molecules can be used to treat disorders associated with abnormal or aberrant metabolism or function of cells in the tissues in which they are expressed. Tissues in which TANGO 275 is expressed 20 include, for example, pancreas, adrenal medulla, adrenal cortex, kidney, thyroid, testis, stomach, heart, brain, liver, placenta, lung, skeletal muscle, or small intestine.

As TANGO 275 exhibits expression in the heart, TANGO 275 polypeptides, nucleic acids, or modulators thereof can be used to treat heart and cardiovascular disorders, such as ischemic heart disease (*e.g.*, angina pectoris, 25 myocardial infarction, and chronic ischemic heart disease), hypertensive heart disease, pulmonary heart disease, valvular heart disease (*e.g.*, rheumatic fever and rheumatic heart disease, endocarditis, mitral valve prolapse, and aortic valve stenosis), congenital heart disease (*e.g.*, valvular and vascular obstructive lesions, atrial or ventricular septal defect, and patent ductus arteriosus), or myocardial disease (*e.g.*, 30 myocarditis, congestive cardiomyopathy, and hypertrophic cardiomyopathy). Disorders of the vasculature that can be treated or prevented according to the methods

of the invention include atheroma, tumor angiogenesis, wound healing, diabetic retinopathy, hemangioma, psoriasis, and restenosis, *e.g.*, restenosis resulting from balloon angioplasty.

In another example, TANGO 275 polypeptides, nucleic acids, or
5 modulators thereof can be used to treat disorders of the brain, such as cerebral edema, hydrocephalus, brain herniations, iatrogenic disease (due to, *e.g.*, infection, toxins, or drugs), inflammations (*e.g.*, bacterial and viral meningitis, encephalitis, and cerebral toxoplasmosis), cerebrovascular diseases (*e.g.*, hypoxia, ischemia, and infarction, intracranial hemorrhage and vascular malformations, and hypertensive
10 encephalopathy), and tumors (*e.g.*, neuroglial tumors, neuronal tumors, tumors of pineal cells, meningeal tumors, primary and secondary lymphomas, intracranial tumors, and medulloblastoma), and to treat injury or trauma to the brain (*e.g.*, spinal cord injuries, infarction, infection, malignancy, exposure to toxic agents, nutritional deficiency, paraneoplastic syndromes), degenerative nerve diseases (including but not
15 limited to Alzheimer's disease, Parkinson's disease, Huntington's Chorea, Gilles de la Tourette's syndrome, amyotrophic lateral sclerosis, progressive supra-nuclear palsy, and other dementias), and neuropsychiatric disorders (including schizophrenia, schizoaffective disorder, attention deficit disorder, dysthymic disorder, major depressive disorder, mania, obsessive-compulsive disorder, psychoactive substance
20 use disorders, anxiety, panic disorder, as well as bipolar affective disorder, *e.g.*, severe bipolar affective disorder, bipolar affective disorder with hypomania and major depression).

In another example, TANGO 275 polypeptides, nucleic acids, or
modulators thereof can be used to treat placental disorders, such as toxemia of
25 pregnancy (*e.g.*, preeclampsia and eclampsia), placentitis, or spontaneous abortion.

In another example, TANGO 275 polypeptides, nucleic acids, or
modulators thereof can be used to treat pulmonary (lung) disorders, such as
atelectasis, cystic fibrosis, rheumatoid lung disease, pulmonary congestion or edema,
chronic obstructive airway disease (*e.g.*, emphysema, chronic bronchitis, bronchial
30 asthma, and bronchiectasis), diffuse interstitial diseases (*e.g.*, sarcoidosis, pneumoconiosis, hypersensitivity pneumonitis, Goodpasture's syndrome, idiopathic

pulmonary hemosiderosis, pulmonary alveolar proteinosis, desquamative interstitial pneumonitis, chronic interstitial pneumonia, fibrosing alveolitis, hamman-rich syndrome, pulmonary eosinophilia, diffuse interstitial fibrosis, Wegener's granulomatosis, lymphomatoid granulomatosis, and lipid pneumonia), or tumors (*e.g.*,
5 bronchogenic carcinoma, bronchioloalveolar carcinoma, bronchial carcinoid, hamartoma, and mesenchymal tumors).

In another example, TANGO 275 polypeptides, nucleic acids, or modulators thereof can be used to treat hepatic disorders, such as jaundice, hepatic failure, liver cysts, chronic liver disease, hereditary hyperbilirubinemias (*e.g.*, Gilbert's
10 syndrome, Crigler-Najjar syndromes and Dubin-Johnson and Rotor's syndromes), hepatic circulatory disorders (*e.g.*, hepatic vein thrombosis and portal vein obstruction and thrombosis) hepatitis (*e.g.*, chronic active hepatitis, acute viral hepatitis, and toxic and drug-induced hepatitis) cirrhosis (*e.g.*, alcoholic cirrhosis, biliary cirrhosis, and hemochromatosis), or malignant tumors (*e.g.*, primary carcinoma, hepatoblastoma,
15 and angiosarcoma).

In another example, TANGO 275 polypeptides, nucleic acids, or modulators thereof can be used to treat disorders of skeletal muscle, such as muscular dystrophy (*e.g.*, Duchenne muscular dystrophy, Becker Muscular Dystrophy, Emery-Dreifuss muscular dystrophy, Limb-Girdle muscular dystrophy, Facioscapulohumeral
20 muscular dystrophy, myotonic dystrophy, oculopharyngeal muscular dystrophy, distal muscular dystrophy, and congenital muscular dystrophy), motor neuron diseases (*e.g.*, amyotrophic lateral sclerosis, infantile progressive spinal muscular atrophy, intermediate spinal muscular atrophy, spinal bulbar muscular atrophy, and adult spinal muscular atrophy), myopathies (*e.g.*, inflammatory myopathies (*e.g.*, dermatomyositis
25 and polymyositis), myotonia congenita, paramyotonia congenita, central core disease, nemaline myopathy, myotubular myopathy, and periodic paralysis), and metabolic diseases of muscle (*e.g.*, phosphorylase deficiency, acid maltase deficiency, phosphofructokinase deficiency, Debrancher enzyme deficiency, mitochondrial myopathy, carnitine deficiency, carnitine palmityl transferase deficiency,
30 phosphoglycerate kinase deficiency, phosphoglycerate mutase deficiency, lactate dehydrogenase deficiency, and myoadenylate deaminase deficiency).

In another example, TANGO 275 polypeptides, nucleic acids, or modulators thereof can be used to treat renal disorders, such as glomerular diseases (*e.g.*, acute and chronic glomerulonephritis, rapidly progressive glomerulonephritis, nephrotic syndrome, focal proliferative glomerulonephritis, glomerular lesions associated with systemic disease, such as systemic lupus erythematosus, Goodpasture's syndrome, multiple myeloma, diabetes, neoplasia, sickle cell disease, and chronic inflammatory diseases), tubular diseases (*e.g.*, acute tubular necrosis and acute renal failure, polycystic renal disease, medullary sponge kidney, medullary cystic disease, nephrogenic diabetes, and renal tubular acidosis), tubulointerstitial diseases (*e.g.*, pyelonephritis, drug and toxin induced tubulointerstitial nephritis, hypercalcemic nephropathy, and hypokalemic nephropathy) acute and rapidly progressive renal failure, chronic renal failure, nephrolithiasis, vascular diseases (*e.g.*, hypertension and nephrosclerosis, microangiopathic hemolytic anemia, atheroembolic renal disease, diffuse cortical necrosis, and renal infarcts), or tumors (*e.g.*, renal cell carcinoma and nephroblastoma).

In another example, TANGO 275 polypeptides, nucleic acids, or modulators thereof can be used to treat pancreatic disorders, such as pancreatitis (*e.g.*, acute hemorrhagic pancreatitis and chronic pancreatitis), pancreatic cysts (*e.g.*, congenital cysts, pseudocysts, and benign or malignant neoplastic cysts), pancreatic tumors (*e.g.*, pancreatic carcinoma and adenoma), diabetes mellitus (*e.g.*, insulin- and non-insulin-dependent types, impaired glucose tolerance, and gestational diabetes), or islet cell tumors (*e.g.*, insulinomas, adenomas, Zollinger-Ellison syndrome, glucagonomas, and somatostatinoma).

TANGO 275 includes an EGF-like domain. Proteins having such domains play a role in a wide variety of biological processes, including cholesterol uptake, blood coagulation, and specification of cell fate. Thus, TANGO 275 polypeptides, nucleic acids, and modulators thereof can be used to modulate these processes. TANGO 275 polypeptides, nucleic acids, and modulators thereof can be used to modulate cell proliferation, morphogenesis, tissue repair and renewal, terminal differentiation, cell survival, and cell migration. They can be used to treat cancer, promote wound healing

(*e.g.*, of the skin, cornea, or mucosa), and modulate an allergic or inflammatory response.

TANGO 275 includes a TB domain. Proteins having this domain are commonly associated with extracellular matrix fibrils. TANGO 275 polypeptides,
5 nucleic acids, and modulators thereof can be used to modulate matrix formation and degradation and to treat disorders of the connective tissue, *e.g.*, Marfan syndrome.

As a transforming growth factor- β binding protein, TANGO 275 can interact with transforming growth factor- β (TGF- β). In general, transforming growth factor- β binding proteins (LTBP) bind to TGF- β to form latent growth factor
10 complexes (large latent complexes). LTBP are important regulators of TGF- β activity. LTBP are thought to facilitate the normal assembly and secretion of large latent complexes, target latent TGF- β to certain connective tissues, modulate the activity of large latent complexes, and target latent TGF- β to the cell surface. Given that TANGO 275 can modulate TGF- β activity, TANGO 275 polypeptides, nucleic
15 acids, and modulators of TANGO 275 expression or activity can be used to treat connective tissue and bone disorders such as bone fracture, osteoporosis, and osteogenesis imperfecta. In addition, such compounds can be used to promote bone repair, promote bone regeneration, and improve bone implant bonding. Thus, TANGO 275 polypeptides, nucleic acids, and modulators thereof can be used to
20 modulate various aspects of bone repair and regeneration, including, *e.g.*, clot formation, clot dissolution, removal of damaged tissue, growth of granulation tissue, cartilage growth and turnover, formation of callus tissue, remodeling, formation of trabecular bone, and formation of cortical bone.

Further, in light of TANGO 275's pattern of expression in humans,
25 TANGO 275 expression can be utilized as a marker for specific tissues (*e.g.*, cardiovascular tissue such as the heart) and/or cells (*e.g.*, cardiac) in which TANGO 275 is expressed. TANGO 275 nucleic acids can also be utilized for chromosomal mapping.

TANGO 300

A cDNA encoding human TANGO 300 was identified by analyzing the sequences of clones present in a human fetal lung cDNA library.

This analysis led to the identification of a sequence encoding full-length
5 human TANGO 300. The cDNA of this clone is 1332 nucleotides long (Figure 17A-17B; SEQ ID NO:13). The 1083 nucleotide open reading frame of this cDNA, nucleotide 31 to nucleotide 1113 of SEQ ID NO:13 (SEQ ID NO:15), encodes a 361 amino acid protein (Figure 17A-17B; SEQ ID NO:14).

The signal peptide prediction program SIGNALP (Nielsen et al., 1997,
10 *Protein Engineering* 10:1-6) predicted that human TANGO 300 includes a 20 amino acid signal peptide (amino acid 1 to about amino acid 20 of SEQ ID NO:14; SEQ ID NO:62) preceding the mature human TANGO 300 protein (corresponding to about amino acid 21 to amino acid 361 of SEQ ID NO:14; SEQ ID NO:63).

Human TANGO 300 is a transmembrane protein having an extracellular
15 domain which extends from about amino acid 21 to about amino acid 304 of SEQ ID NO:14 (SEQ ID NO:85), a transmembrane domain which extends from about amino acid 305 to about amino acid 321 of SEQ ID NO:14 (SEQ ID NO:86), and a cytoplasmic domain which extends from about amino acid 322 to amino acid 361 of SEQ ID NO:14 (SEQ ID NO:87).

20 Alternatively, in another embodiment, a human TANGO 300 protein contains an extracellular domain at amino acid residues 322 to amino acid 361 of SEQ ID NO:14 (SEQ ID NO:87), transmembrane domains at amino acid residues 305 to about amino acid 321 of SEQ ID NO:14 (SEQ ID NO:86), and a cytoplasmic domain at amino acid 21 to about amino acid 304 of SEQ ID NO:14 (SEQ ID NO:85).

25 Human TANGO 300 that has not been post-translationally modified is predicted to have a molecular weight of 40.6 kDa prior to cleavage of its signal peptide and a molecular weight of 38.5 kDa subsequent to cleavage of its signal peptide.

Within human TANGO 300, protein kinase C phosphorylation sites are
30 present at amino acids 74 to 76, 89 to 91, 307 to 309, and 359 to 361 of SEQ ID NO:14. Casein kinase II phosphorylation sites are present at amino acids 34 to 37, 41

to 44, 74 to 77, 153 to 156, and 169 to 172 of SEQ ID NO:14. Tyrosine kinase phosphorylation sites are present at amino acids 111 to 117 and 236 to 243 of SEQ ID NO:14. N-myristylation sites are present at amino acids 25 to 30 and 170 to 175 of SEQ ID NO:14.

5 Clone AthX672i5, which encodes human TANGO 300, was deposited as EpT300 with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2236) on June 30, 1999 and assigned Accession Number PTA-293. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the
10 Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 18 depicts a hydropathy plot of human TANGO 300. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions
15 are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 18 indicates that human TANGO 300 has a signal peptide at its amino terminus and an internal hydrophobic region, suggesting that human TANGO 300 is a transmembrane protein.

20 A clone, jthub009c07, containing murine TANGO 300 was also identified. The cDNA of this clone is 1400 nucleotides long (Figures 19A-19C; SEQ ID NO:16). The 1155 nucleotide open reading frame of this cDNA, nucleotide 41 to nucleotide 1195 of SEQ ID NO:16 (SEQ ID NO:18), encodes a 385 amino acid protein (Figures 19A-19C; SEQ ID NO:17).

25 The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that murine TANGO 300 includes a 19 amino acid signal peptide (amino acid 1 to about amino acid 19 of SEQ ID NO:16; SEQ ID NO:64) preceding the mature murine TANGO 300 protein (corresponding to about amino acid 20 to amino acid 385 of SEQ ID NO:16; SEQ ID NO:65).

30 Murine TANGO 300 is a transmembrane protein having an extracellular domain which extends from about amino acid 20 to about amino acid 318 of SEQ ID

NO:16 (SEQ ID NO:88), a transmembrane domain which extends from about amino acid 319 to about amino acid 335 of SEQ ID NO:16 (SEQ ID NO:89), and a cytoplasmic domain which extends from about amino acid 336 to amino acid 385 of SEQ ID NO:16 (SEQ ID NO:90).

5 Alternatively, in another embodiment, a murine TANGO 300 protein contains an extracellular domain at amino acid residues 336 to amino acid 385 of SEQ ID NO:16 (SEQ ID NO:90), transmembrane domains at amino acid residues 319 to about amino acid 335 of SEQ ID NO:16 (SEQ ID NO:89), and a cytoplasmic domain at amino acid 20 to about amino acid 318 of SEQ ID NO:16 (SEQ ID NO:88).

10 Murine TANGO 300 that has not been post-translationally modified is predicted to have a molecular weight of 43.1 kDa prior to cleavage of its signal peptide and a molecular weight of 41.0 kDa subsequent to cleavage of its signal peptide.

Within murine TANGO 300, protein kinase C phosphorylation sites are
 15 present at amino acids 85 to 87 and 378 to 380 of SEQ ID NO:17. Casein kinase II phosphorylation sites are present at amino acids 22 to 25, 37 to 40, 149 to 152, 165 to 168 and 287 to 290 of SEQ ID NO:17. A tyrosine kinase phosphorylation site is present at amino acids 107 to 113 of SEQ ID NO:17. N-myristylation sites are present at amino acids 29 to 34, 89 to 94, 166 to 171 and 207 to 212 of SEQ ID
 20 NO:17. A N-glycosylation site is present at amino acids 136 to 139 of SEQ ID NO:17.

Figure 20 depicts a hydropathy plot of murine TANGO 300. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short
 25 vertical lines just below the hydropathy trace. The hydropathy plot of Figure 20 indicates that murine TANGO 300 has a signal peptide at its amino terminus and an internal hydrophobic region, suggesting that murine TANGO 300 is a transmembrane protein.

Figures 21A-21B depict an alignment of the ORF nucleotide sequence of
 30 human TANGO 300 (SEQ ID NO:15) and the ORF nucleotide sequence of murine TANGO 300 (SEQ ID NO:18). This alignment was created using BESTFIT

(BLOSUM 62 scoring matrix; gap open penalty of 12; frame shift penalty of 5; gap extend penalty of 4). In this alignment, the sequences are 77.7% identical. Figure 22 depicts an alignment of the amino acid sequence of human TANGO 300 (SEQ ID NO:14) and the amino acid sequence of murine TANGO 300 (SEQ ID NO:17). This alignment was created using BESTFIT (BLOSUM 62 scoring matrix; gap open penalty of 12; frame shift penalty of 5; gap extend penalty of 4). In this alignment, the sequences are 69.6% identical. The full length nucleotide sequences of human TANGO 300 and murine TANGO 300 display 75.8% identity.

10 Use of TANGO 300 Nucleic Acids, Polypeptides, and Modulators Thereof

TANGO 300 polypeptides, nucleic acids, and modulators thereof can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues in which they are expressed.

Further, in light of TANGO 300's presence in a fetal lung cDNA library, TANGO 300 expression can be utilized as a marker for specific tissues (*e.g.*, lung) and/or cells (*e.g.*, pulmonary) in which TANGO 300 is expressed. TANGO 300 nucleic acids can also be utilized for chromosomal mapping.

MANGO 245

20 A cDNA encoding MANGO 245 was identified by analyzing the sequences of clones present in a human adult brain cDNA library.

This analysis led to the identification of a clone, Alhbab165e5, encoding full-length human MANGO 245. The cDNA of this clone is 1356 nucleotides long (Figures 23A-23B; SEQ ID NO:19). The 1044 nucleotide open reading frame of this cDNA, nucleotide 105 to nucleotide 1148 of SEQ ID NO:19 (SEQ ID NO:21), encodes a 348 amino acid protein (Figures 23A-23B; SEQ ID NO:20).

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that human MANGO 245 includes a 16 amino acid signal peptide (amino acid 1 to about amino acid 16 of SEQ ID NO:20; SEQ ID NO:66) preceding the mature human MANGO 245 protein (corresponding to about amino acid 17 to amino acid 348 of SEQ ID NO:20; SEQ ID NO:67).

Human MANGO 245 is a transmembrane protein having an extracellular domain which extends from about amino acid 17 to about amino acid 141 of SEQ ID NO:20 (SEQ ID NO:78), a transmembrane domain which extends from about amino acid 142 to about amino acid 158 of SEQ ID NO:20 (SEQ ID NO:79), and a
5 cytoplasmic domain which extends from about amino acid 159 to amino acid 348 of SEQ ID NO:20 (SEQ ID NO:80).

Alternatively, in another embodiment, a murine TANGO 300 protein contains an extracellular domain at amino acid residues 159 to amino acid 348 of SEQ ID NO:20 (SEQ ID NO:80), transmembrane domains at amino acid residues 142 to
10 about amino acid 158 of SEQ ID NO:20 (SEQ ID NO:79), and a cytoplasmic domain at amino acid 17 to about amino acid 141 of SEQ ID NO:20 (SEQ ID NO:78).

Human MANGO 245 that has not been post-translationally modified is predicted to have a molecular weight of 37.9 kDa prior to cleavage of its signal peptide and a molecular weight of 36.3 kDa subsequent to cleavage of its signal
15 peptide.

Human MANGO 245 includes CIq domains at amino acids 31 to 156 of SEQ ID NO:20 (SEQ ID NO:70) and amino acids 178 to 294 of SEQ ID NO:20 (SEQ ID NO:71). Figure 27 depicts alignments of the CIq domains of human MANGO 245 with a consensus hidden Markov model CIq domain (SEQ ID NO:72).

20 Within MANGO 245, protein kinase C phosphorylation sites are present at amino acids 244 to 246 and 264 to 266 of SEQ ID NO:20. Casein kinase II phosphorylation sites are present at amino acids 38 to 41 and 298 to 301 of SEQ ID NO:20. N-myristylation sites are present at amino acids 66 to 71, 113 to 118, 145 to 150, 219 to 224, and 295 to 300 of SEQ ID NO:20.

25 Clone Alhbab165e5, which encodes human MANGO 245, was deposited as EpM245 with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on April 21, 1999 and assigned Accession Number 207223. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the
30 Purposes of Patent Procedure. This deposit was made merely as a convenience for

those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 24 depicts a hydropathy plot of human MANGO 245. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 24 indicates that human MANGO 245 has a signal peptide at its amino terminus and an internal hydrophobic region, suggesting that human MANGO 245 is a transmembrane protein.

10 Northern blot analysis of human MANGO 245 expression revealed that human MANGO 245 is expressed at a relatively high level in the cerebellum, frontal lobe, and putamen; at a moderate level in the cerebral cortex, the medulla, occipital lobe, and temporal lobe; and a relatively low level in the spinal cord. Additional Northern blot analysis revealed the human MANGO 245 is expressed in amygdala, 15 caudate nucleus, hippocampus, brain, substantia nigra, and subthalamic nucleus.

A cDNA encoding monkey MANGO 245 was identified by analyzing the sequences of clones present in a monkey cDNA library.

This analysis led to the identification of a clone, Alkbd75h1, encoding full-length monkey MANGO 245. The cDNA of this clone is 1416 nucleotides long 20 (Figures 25A-25B; SEQ ID NO:22). The 987 nucleotide open reading frame of this cDNA, nucleotide 250 to nucleotide 1236 of SEQ ID NO:22 (SEQ ID NO:24), encodes a 329 amino acid protein (Figures 25A-25B; SEQ ID NO:23).

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that monkey MANGO 245 includes a 16 25 amino acid signal peptide (amino acid 1 to about amino acid 16 of SEQ ID NO:23; SEQ ID NO:68) preceding the mature monkey MANGO 245 protein (corresponding to about amino acid 17 to amino acid 329 of SEQ ID NO:23; SEQ ID NO:69).

Monkey MANGO 245 that has not been post-translationally modified is predicted to have a molecular weight of 35.2 kDa prior to cleavage of its signal 30 peptide and a molecular weight of 33.6 kDa subsequent to cleavage of its signal peptide.

Monkey MANGO 245 includes CIq domains at amino acids 31 to 156 of SEQ ID NO:23 (SEQ ID NO:73) and amino acids 178 to 311 of SEQ ID NO:23 (SEQ ID NO:74). Figure 28 depicts alignments of the CIq domains of monkey MANGO 245 with a consensus hidden Markov model CIq domain (SEQ ID NO:72).

5 Figure 26 depicts an alignment of the amino acid sequence of human MANGO 245 (SEQ ID NO:20) and the amino acid sequence of monkey MANGO 245 (SEQ ID NO:23). This alignment was created using ALIGN (version 2.0; PAM120 scoring matrix; gap length penalty of 12; gap penalty of 4). In this alignment, the sequences are 84.8% identical overall.

10 Clone Alkbd75h1, which encodes monkey MANGO 245, was deposited as EpK245 with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on June 18, 1999 and assigned Accession Number PTA-248. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the
15 Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

 In addition, a murine MANGO 245 was identified. The cDNA of this clone is 625 nucleotides long (Figure 29; SEQ ID NO:25). The open reading frame of
20 this cDNA begins at nucleotide 29 of SEQ ID NO:25. Murine MANGO 245 includes a CIq domain at amino acids 30 to 155 of SEQ ID NO: 91 (SEQ ID NO:93).

 Within murine MANGO 245, protein kinase C phosphorylation sites are present at amino acids 64 to 66 and 178 to 180 of SEQ ID NO:91. N-myristylation sites are present at amino acids 112 to 117 and 144 to 149 of SEQ ID NO:91. A
25 casein kinase II phosphorylation site is present at amino acids 37 to 40 of SEQ ID NO:91. An N-glycosylation site is present at amino acids 88 to 91 of SEQ ID NO:91.

 Figures 30A-30B depict an alignment of 697 of the 1356 nucleotides of the human MANGO 245 sequence (nucleotide 51 to nucleotide 748 of SEQ ID NO:19) with the nucleotide sequence of murine MANGO 245 (SEQ ID NO:25). This
30 alignment was created using BESTFIT (BLOSUM 62 scoring matrix; gap open

penalty of 12; frame shift penalty of 5; gap extend penalty of 4). In this alignment, the sequences are 89.6% identical overall.

Use of MANGO 245 Nucleic Acids, Polypeptides, and Modulators Thereof

5 MANGO 245 polypeptides, nucleic acids, and modulators thereof can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues in which they are expressed. MANGO 245 is expressed in the brain and central nervous system. Thus, MANGO 245 polypeptides, nucleic acids, and modulators thereof can be used to treat CNS disorders such as Alzheimer's
10 disease, senile dementia, Huntington's disease, amyotrophic lateral sclerosis, and Parkinson's disease, as well as Gilles de la Tourette's syndrome, autonomic function disorders such as hypertension and sleep disorders, and neuropsychiatric disorders that include, but are not limited to schizophrenia, schizoaffective disorder, attention deficit disorder, dysthymic disorder, major depressive disorder, mania, obsessive-
15 compulsive disorder, psychoactive substance use disorders, anxiety, panic disorder, as well as bipolar affective disorder, *e.g.*, severe bipolar affective (mood) disorder (BP-I), bipolar affective (mood) disorder with hypomania and major depression (BP-II).

MANGO 245 includes a CIq domain. Known proteins having this domain play a role complement activation and autoimmune disorders. The CIq domain is also
20 found in collagens and collagen-like molecules. MANGO 245 polypeptides, nucleic acids, and modulators thereof can be used to treat disorders of collagen assembly and degradation.

Further, in light of MANGO 245's pattern of expression in humans, MANGO 245 expression can be utilized as a marker for specific tissues (*e.g.*, brain)
25 and/or cells (*e.g.*, cerebellum, frontal lobe, or putamen) in which MANGO 245 is expressed. MANGO 245 nucleic acids can also be utilized for chromosomal mapping.

Tables 1 and 2 below provide summaries of TANGO 244, TANGO 246, TANGO 275, TANGO 300 and MANGO 245 sequence and protein domain information.

5 TABLE 1: Summary of Sequence Information for TANGO 244, TANGO 246, TANGO 275, TANGO 300 and MANGO 245

	Gene	cDNA	ORF	Polypeptide	Figure	ATCC® Accession Number
10	TANGO 244 Human	SEQ ID NO:1	SEQ ID NO:3	SEQ ID NO:2	Fig. 1	207223
	TANGO 246 Human	SEQ ID NO:4	SEQ ID NO:6	SEQ ID NO:5	Fig. 5A-5B	207223
	TANGO 275 Human	SEQ ID NO:7	SEQ ID NO:9	SEQ ID NO:8	Fig. 9A-9D	207220
15	TANGO 245 Mouse	SEQ ID NO:10	SEQ ID NO:12	SEQ ID NO:11	Fig. 16A-16G	---
	TANGO 300 Human	SEQ ID NO:13	SEQ ID NO:15	SEQ ID NO:14	Fig. 17A-17B	PTA-293
20	TANGO 300 Mouse	SEQ ID NO:16	SEQ ID NO:18	SEQ ID NO:17	Fig. 19A-19C	---
	MANGO 245 Human	SEQ ID NO:19	SEQ ID NO:21	SEQ ID NO:20	Fig. 23	207223
	MANGO 245 Monkey	SEQ ID NO:22	SEQ ID NO:24	SEQ ID NO:23	Fig. 25A-25B	PTA-248
25	MANGO 245 Mouse	SEQ ID NO:25	SEQ ID NO:92	SEQ ID NO:91	Fig. 29	---

TABLE 2: Summary of Protein Domains of TANGO 244, TANGO 246, TANGO 275, and TANGO 300 and MANGO 245

	Protein	Signal Peptide	Mature Protein	Extracellular Domain	Transmembrane Domain	Cytoplasmic Domain
5	TANGO 244 Human	AA 1-26 of SEQ ID NO:2 SEQ ID NO:26	AA 27-162 of SEQ ID NO:2 SEQ ID NO:27	AA 27-119 of SEQ ID NO:2 SEQ ID NO:75	AA 120-142 of SEQ ID NO:2 SEQ ID NO:76	AA 143-162 of SEQ ID NO:2 SEQ ID NO:77
	TANGO 246 Human	—	—	AA 1-305 of SEQ ID NO:5 SEQ ID NO:57	AA 306-323 of SEQ ID NO:5 SEQ ID NO:58	AA 324-329 of SEQ ID NO:5 SEQ ID NO:59
10	TANGO 275 Human	AA 1-29 of SEQ ID NO:8 SEQ ID NO:60	AA 30-1303 of SEQ ID NO:8 SEQ ID NO:61	---	---	---
	TANGO 300 Human	AA 1-20 of SEQ ID NO:14 SEQ ID NO:62	AA 21-361 of SEQ ID NO:14 SEQ ID NO:63	AA 21-304 of SEQ ID NO:14 SEQ ID NO:85	AA 305-321 of SEQ ID NO:14 SEQ ID NO:86	AA 322-361 of SEQ ID NO:14 SEQ ID NO:87
	TANGO 300 Mouse	AA 1-19 of SEQ ID NO:17 SEQ ID NO:64	AA 20-385 of SEQ ID NO:17 SEQ ID NO:65	AA 20-318 of SEQ ID NO:17 SEQ ID NO:88	AA 319-335 of SEQ ID NO:17 SEQ ID NO:89	AA 336-385 of SEQ ID NO:17 SEQ ID NO:90
15	MANGO 245 Human	AA 1-16 of SEQ ID NO:20 SEQ ID NO:66	AA 17-348 of SEQ ID NO:20 SEQ ID NO:67	AA 17-141 of SEQ ID NO:20 SEQ ID NO:78	AA 142-158 of SEQ ID NO:20 SEQ ID NO:79	AA 159-348 of SEQ ID NO:20 SEQ ID NO:80
	MANGO 245 Monkey	AA 1-16 of SEQ ID NO:23 SEQ ID NO:68	AA 17-329 of SEQ ID NO:23 SEQ ID NO:69	---	---	---
20	MANGO 245 Mouse	AA 1-16 of SEQ ID NO:91	AA 17-199 of SEQ ID NO:91	---	---	---

Deposit Information

Clone Athua62f9, which encodes human TANGO 244, was deposited, as part of a composite deposit, as EpT244 with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on April 21, 1999 and assigned Accession Number 207223.

Clone Athsa34d2, which encodes human TANGO 246, was deposited, as part of a composite deposit, as EpT246 with the American Type Culture Collection

(ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on April 21, 1999 and assigned Accession Number 207223.

Clone Athbb19d1, which encodes human TANGO 275, was deposited, as part of a composite deposit, as EpT275 with the American Type Culture Collection

5 (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on April 21, 1999 and assigned Accession Number 207220.

Clone AthX672i5, which encodes human TANGO 300, was deposited as EpT300 with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2236) on June 30, 1999, and assigned Accession

10 Number PTA-293.

Clone Alhbab165e5, which encodes human MANGO 245, was deposited, as part of a composite deposit, as EpM245 with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on April 21, 1999 and assigned Accession Number 207223.

15 Clone Alkbd75h1, which encodes monkey MANGO 245, was deposited as EpK245 with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on June 18, 1999 and assigned Accession Number PTA-248.

The clones containing cDNA molecules encoding human TANGO 244, human
20 TANGO 246, and human MANGO 245 were deposited with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on April 21, 1999 as Accession Number 207223, as part of a composite deposit representing a mixture of five strains, each carrying one recombinant plasmid harboring a particular cDNA clone.

25 To distinguish the strains and isolate a strain harboring a particular cDNA clone, an aliquot of the mixture can be streaked out to single colonies on nutrient medium (*e.g.*, LB plates) supplemented with 100µg/ml ampicillin, single colonies grown, and then plasmid DNA extracted using a standard miniprep procedure. Next, a sample of the DNA miniprep can be digested with a combination of the
30 restriction enzymes *Sal* I, *Not* I, and *Sac* II and the resultant products resolved on a 0.8% agarose gel using standard DNA electrophoresis conditions. The digest liberates

fragments as follows: human TANGO 244 (1.5 kB), human TANGO 246 (2.0 kB), human MANGO 245 (0.7 kB and 0.65 kB; human MANGO 245 has a *Sac* II cut site at about bp 693). The identity of the strains can be inferred from the fragments liberated.

5 Various aspects of the invention are described in further detail in the following subsections.

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that
10 encode a polypeptide of the invention or a biologically active portion thereof, as well as nucleic acid molecules sufficient for use as hybridization probes to identify nucleic acid molecules encoding a polypeptide of the invention and fragments of such nucleic acid molecules suitable for use as PCR primers for the amplification or mutation of nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended
15 to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other
20 nucleic acid molecules which are present in the natural source of the nucleic acid molecule. Preferably, an "isolated" nucleic acid molecule is free of sequences (preferably protein encoding sequences) which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. In other embodiments, the
25 "isolated" nucleic acid is free of intron sequences. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kB, 4 kB, 3 kB, 2 kB, 1 kB, 0.5 kB or 0.1 kB of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule,
30 can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other

chemicals when chemically synthesized. In one embodiment, the nucleic acid molecules of the invention comprise a contiguous open reading frame encoding a polypeptide of the invention.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule
5 having the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, or a complement thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequences of SEQ ID NO: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25 as a hybridization probe, nucleic acid molecules of the
10 invention can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook et al., eds., *Molecular Cloning: A Laboratory Manual, 2nd ed.*, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

A nucleic acid molecule of the invention can be amplified using cDNA,
15 mRNA or genomic DNA as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to all or a portion of a nucleic acid molecule of the invention can be prepared by standard synthetic techniques, *e.g.*,
20 using an automated DNA synthesizer.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, or a portion thereof. A nucleic acid molecule which is complementary to a given
25 nucleotide sequence is one which is sufficiently complementary to the given nucleotide sequence that it can hybridize to the given nucleotide sequence thereby forming a stable duplex.

Moreover, a nucleic acid molecule of the invention can comprise only a portion of a nucleic acid sequence encoding a full length polypeptide of the invention
30 for example, a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a polypeptide of the invention. The

nucleotide sequence determined from the cloning one gene allows for the generation of probes and primers designed for use in identifying and/or cloning homologues in other cell types, *e.g.*, from other tissues, as well as homologues from other mammals. The probe/primer typically comprises substantially purified oligonucleotide. The
5 oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 50, 75, 100, 125, 150, 175, 200, 250, 300, 350 or 400 consecutive nucleotides of the sense or anti-sense sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, or of a naturally occurring mutant of SEQ ID NO:1, 3, 4, 6, 7, 9,
10 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25.

Probes based on the sequence of a nucleic acid molecule of the invention can be used to detect transcripts or genomic sequences encoding the same protein molecule encoded by a selected nucleic acid molecule. The probe comprises a label group attached thereto, *e.g.*, a radioisotope, a fluorescent compound, an enzyme, or an
15 enzyme co-factor. Such probes can be used as part of a diagnostic test kit for identifying cells or tissues which mis-express the protein, such as by measuring levels of a nucleic acid molecule encoding the protein in a sample of cells from a subject, *e.g.*, detecting mRNA levels or determining whether a gene encoding the protein has been mutated or deleted.

20 A nucleic acid fragment encoding a biologically active portion of a polypeptide of the invention can be prepared by isolating a portion of any of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, expressing the encoded portion of the polypeptide protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the polypeptide.

25 The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25.

30 In addition to the nucleotide sequences of SEQ ID NO: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, it will be appreciated by those skilled in the art that

DNA sequence polymorphisms that lead to changes in the amino acid sequence may exist within a population (e.g., the human population). Such genetic polymorphisms may exist among individuals within a population due to natural allelic variation. An allele is one of a group of genes which occur alternatively at a given genetic locus.

- 5 As used herein, the phrase "allelic variant" refers to a nucleotide sequence which occurs at a given locus or to a polypeptide encoded by the nucleotide sequence. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a polypeptide of the invention. Such natural allelic variations can typically result in 1-5% variance in the nucleotide
- 10 sequence of a given gene. Alternative alleles can be identified by sequencing the gene of interest in a number of different individuals. This can be readily carried out by using hybridization probes to identify the same genetic locus in a variety of individuals. Any and all such nucleotide variations and resulting amino acid polymorphisms or variations that are the result of natural allelic variation and that do
- 15 not alter the functional activity are intended to be within the scope of the invention.

- Moreover, nucleic acid molecules encoding proteins of the invention from other species (homologues), which have a nucleotide sequence which differs from that of the human protein described herein are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and
- 20 homologues of a cDNA of the invention can be isolated based on their identity to the human nucleic acid molecule disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a cDNA encoding a soluble form of a membrane-bound protein of the invention isolated based on its hybridization to a
- 25 nucleic acid molecule encoding all or part of the membrane-bound form. Likewise, a cDNA encoding a membrane-bound form can be isolated based on its hybridization to a nucleic acid molecule encoding all or part of the soluble form.

- Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800,
- 30 900, 1000, or 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3200, 3400, 3600, 3800, 4000, or 4200) nucleotides in length and hybridizes under stringent

conditions to the nucleic acid molecule comprising the nucleotide sequence, preferably the coding sequence, of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, or a complement thereof.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% (65%, 70%, preferably 75% or more) identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6, which describes aqueous and non-aqueous methods, either of which can be used. Another preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 2.0 X SSC at 50° C. (low stringency) or 0.2 X SSC, 0.1% SDS at 50-65°C (high stringency). Another preferred example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Preferably, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. Particularly preferred stringency conditions (and the conditions that should be used if the practitioner is uncertain about what conditions should be applied to determine if a molecule is within a hybridization limitation of the invention) are 0.5M Sodium Phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C. In one embodiment, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, or a complement thereof, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a

"naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In addition to naturally-occurring allelic variants of a nucleic acid molecule of the invention sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation thereby leading to changes in the amino acid sequence of the encoded protein, without altering the biological activity of the protein. For example, one can make nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are not conserved or only semi-conserved among homologues of various species may be non-essential for activity and thus would be likely targets for alteration. Alternatively, amino acid residues that are conserved among the homologues of various species (*e.g.*, murine and human) may be essential for activity and thus would not be likely targets for alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding a polypeptide of the invention that contain changes in amino acid residues that are not essential for activity. Such polypeptides differ in amino acid sequence from SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule includes a nucleotide sequence encoding a protein that includes an amino acid sequence that is at least about 45% identical, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91.

An isolated nucleic acid molecule encoding a variant protein can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Briefly, PCR primers are designed that delete the trinucleotide codon of the amino

acid to be changed and replace it with the trinucleotide codon of the amino acid to be included. This primer is used in the PCR amplification of DNA encoding the protein of interest. This fragment is then isolated and inserted into the full length cDNA encoding the protein of interest and expressed recombinantly. The resulting protein

5 now includes the amino acid replacement.

Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. Conservative replacements are those that take place within a family of amino acids that are related in their side chains.

Genetically encoded amino acids are can be divided into four families: (1) acidic =
10 aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) nonpolar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. In similar fashion, the amino acid repertoire can be grouped as (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine histidine, (3) aliphatic = glycine,
15 alanine, valine, leucine, isoleucine, serine, threonine, with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic = phenylalanine, tyrosine, tryptophan; (5) amide = asparagine, glutamine; and (6) sulfur -containing = cysteine and methionine. (See, for example, Biochemistry, 4th ed., Ed. by L. Stryer, WH Freeman and Co.: 1995).

20 Alternatively, mutations can be introduced randomly along all or part of the coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for biological activity to identify mutants that retain activity. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

25 In a preferred embodiment, a mutant polypeptide that is a variant of a polypeptide of the invention can be assayed for: (1) the ability to form protein-protein interactions with proteins in a signaling pathway of the polypeptide of the invention; (2) the ability to bind a ligand of the polypeptide of the invention; or (3) the ability to bind to an intracellular target protein of the polypeptide of the invention. In yet
30 another preferred embodiment, the mutant polypeptide can be assayed for the ability

to modulate cellular proliferation, cellular migration or chemotaxis, or cellular differentiation.

The present invention encompasses antisense nucleic acid molecules, i.e., molecules which are complementary to a sense nucleic acid encoding a polypeptide of the invention, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire coding strand, or to only a portion thereof, *e.g.*, all or part of the protein coding region (or open reading frame). An antisense nucleic acid molecule can be antisense to all or part of a non-coding region of the coding strand of a nucleotide sequence encoding a polypeptide of the invention. The non-coding regions ("5' and 3' untranslated regions") are the 5' and 3' sequences which flank the coding region and are not translated into amino acids.

An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, β -D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, β -D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine,

pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically
5 using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically
10 administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a selected polypeptide of the invention to thereby inhibit expression, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid
15 molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration,
20 antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the
25 antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

An antisense nucleic acid molecule of the invention can be an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units,
30 the strands run parallel to each other (Gaultier et al., 1987, *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-

methyribonucleotide (Inoue et al., 1987, *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al., 1987, *FEBS Lett.* 215:327-330).

The invention also encompasses ribozymes. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded
5 nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes; described in Haselhoff and Gerlach, 1988, *Nature* 334:585-591) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of the protein encoded by the mRNA. A ribozyme having specificity for a nucleic acid molecule encoding a polypeptide of the invention can be
10 designed based upon the nucleotide sequence of a cDNA disclosed herein. For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, an mRNA encoding a polypeptide of the invention can
15 be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel and Szostak, 1993, *Science* 261:1411-1418.

The invention also encompasses nucleic acid molecules which form triple helical structures. For example, expression of a polypeptide of the invention can be inhibited by targeting nucleotide sequences complementary to the regulatory region of
20 the gene encoding the polypeptide (*e.g.*, the promoter and/or enhancer) to form triple helical structures that prevent transcription of the gene in target cells. *See generally* Helene, 1991, *Anticancer Drug Des.* 6(6):569-84; Helene, 1992, *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, 1992, *Bioassays* 14(12):807-15.

In various embodiments, the nucleic acid molecules of the invention can be
25 modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (*see* Hyrup et al., 1996, *Bioorganic & Medicinal Chemistry* 4(1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics,
30 *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The

neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al., 1996, *supra*; Perry-O'Keefe et al., 1996, *Proc. Natl. Acad. Sci. USA*

5 93: 14670-675.

PNAs can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs can also be used, *e.g.*, in the analysis of single base pair mutations
10 in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup, 1996, *supra*; or as probes or primers for DNA sequence and hybridization (Hyrup, 1996, *supra*; Perry-O'Keefe et al., 1996, *Proc. Natl. Acad. Sci. USA* 93: 14670-675).

In another embodiment, PNAs can be modified, *e.g.*, to enhance their stability
15 or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact
20 with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup, 1996, *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, 1996, *supra*, and Finn et al., 1996, *Nucleic Acids*
25 *Res.* 24(17):3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs. Compounds such as 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite can be used as a link between the PNA and the 5' end of DNA (Mag et al., 1989, *Nucleic Acids Res.* 17:5973-88). PNA monomers are then coupled in a
30 stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn et al., 1996, *Nucleic Acids Res.* 24(17):3357-63). Alternatively,

chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser et al., 1975, *Bioorganic Med. Chem. Lett.* 5:1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger et al., 1989, *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre et al., 1987, *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W0 88/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W0 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, e.g., Krol et al., 1988, *Bio/Techniques* 6:958-976) or intercalating agents (see, e.g., Zon, 1988, *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

15 II. Isolated Proteins and Antibodies

One aspect of the invention pertains to isolated proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise antibodies directed against a polypeptide of the invention. In one embodiment, the native polypeptide can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, polypeptides of the invention are produced by recombinant DNA techniques. Alternative to recombinant expression, a polypeptide of the invention can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free of chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, or 5% (by

dry weight) of heterologous protein (also referred to herein as a "contaminating protein"). When the protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When the protein is produced by chemical synthesis, it is preferably substantially free of chemical precursors or other chemicals, i.e., it is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. Accordingly such preparations of the protein have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or compounds other than the polypeptide of interest. The term "pure" or "isolated" as used herein preferably has the same numerical limits as "purified" or "isolated" immediately above. "Isolated" and "purified" do not encompass either natural materials in their native state or natural materials that have been separated into components (*e.g.*, in an acrylamide gel) but not obtained either as pure (*e.g.*, lacking contaminating proteins, or chromatography reagents such as denaturing agents and polymers, *e.g.*, acrylamide or agarose) substances or solutions. In preferred embodiments, purified or isolated preparations will lack any contaminating proteins from the same animal from which the protein is normally produced, as can be accomplished by recombinant expression of, for example, a human protein in a non-human cell.

Biologically active portions of a polypeptide of the invention include polypeptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the protein (*e.g.*, the amino acid sequence shown in any of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91), which include fewer amino acids than the full length protein, and exhibit at least one activity of the corresponding full-length protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the corresponding protein. A biologically active portion of a protein of the invention can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of the native form of a polypeptide of the invention.

Preferred polypeptides have the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91. Other useful proteins are substantially identical (*e.g.*, at least about 45%, preferably 55%, 65%, 75%, 85%, 95%, or 99%) to any of SEQ ID NOs: 2, 5, 8, 11, 14, 17, 20, 23, or 91 and retain the functional activity of the protein of the
5 corresponding naturally-occurring protein yet differ in amino acid sequence due to natural allelic variation or mutagenesis.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for
10 optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity
15 between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = # of identical positions/total # of positions (*e.g.*, overlapping positions) x 100). In one embodiment the two sequences are the same length.

The determination of percent identity between two sequences can be
20 accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul, 1990, *Proc. Natl. Acad. Sci. USA* 87:2264-2268, modified as in Karlin and Altschul, 1993, *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of
25 Altschul, et al., 1990, *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to a protein molecules of
30 the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., 1997, *Nucleic Acids Res.*

25:3389-3402. Alternatively, PSI-Blast can be used to perform an iterated search which detects distant relationships between molecules (*Id.*). When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. *See*

5 <http://www.ncbi.nlm.nih.gov>. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, 1988, *CABIOS* 4:11-17. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a

10 PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, typically exact matches are counted.

15 The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises all or part (preferably biologically active) of a polypeptide of the invention operably linked to a heterologous polypeptide (*i.e.*, a polypeptide other than the same polypeptide of the invention). Within the fusion protein, the term "operably linked" is intended to indicate that the

20 polypeptide of the invention and the heterologous polypeptide are fused in-frame to each other. The heterologous polypeptide can be fused to the N-terminus or C-terminus of the polypeptide of the invention.

One useful fusion protein is a GST fusion protein in which the polypeptide of the invention is fused to the C-terminus of GST sequences. Such fusion proteins can

25 facilitate the purification of a recombinant polypeptide of the invention.

In another embodiment, the fusion protein contains a heterologous signal peptide at its N-terminus. For example, the native signal peptide of a polypeptide of the invention can be removed and replaced with a signal peptide from another protein. For example, the gp67 secretory sequence of the baculovirus envelope protein can be

30 used as a heterologous signal peptide (*Current Protocols in Molecular Biology*, Ausubel et al., eds., John Wiley & Sons, 1992). Other examples of eukaryotic

heterologous signal peptides include the secretory sequences of melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal peptides include the *phoA* secretory signal (Sambrook et al., *supra*) and the protein A secretory signal (Pharmacia Biotech;

5 Piscataway, New Jersey).

In yet another embodiment, the fusion protein is an immunoglobulin fusion protein in which all or part of a polypeptide of the invention is fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions
10 and administered to a subject to inhibit an interaction between a ligand (soluble or membrane-bound) and a protein on the surface of a cell (receptor), to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion protein can be used to affect the bioavailability of a cognate ligand of a polypeptide of the invention. Inhibition of ligand/receptor interaction may be useful therapeutically, both for treating
15 proliferative and differentiative disorders and for modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies directed against a polypeptide of the invention in a subject, to purify ligands and in screening assays to identify molecules which inhibit the interaction of receptors with ligands.

20 Chimeric and fusion proteins of the invention can be produced by standard recombinant DNA techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene
25 fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.*, Ausubel et al., *supra*). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the
30 polypeptide of the invention.

A signal peptide of a polypeptide of the invention (SEQ ID NOs:26, 60, 62, 64, 66, or 68) can be used to facilitate secretion and isolation of the secreted protein or other proteins of interest. Signal peptides are typically characterized by a core of hydrophobic amino acids which are generally cleaved from the mature protein during secretion in one or more cleavage events. Such signal peptides contain processing sites that allow cleavage of the signal peptide from the mature proteins as they pass through the secretory pathway. Thus, the invention pertains to the described polypeptides having a signal peptide, as well as to the signal peptide itself and to the polypeptide in the absence of the signal peptide (i.e., the cleavage products). In one embodiment, a nucleic acid sequence encoding a signal peptide of the invention can be operably linked in an expression vector to a protein of interest, such as a protein which is ordinarily not secreted or is otherwise difficult to isolate. The signal peptide directs secretion of the protein, such as from a eukaryotic host into which the expression vector is transformed, and the signal peptide is subsequently or concurrently cleaved. The protein can then be readily purified from the extracellular medium by art recognized methods. Alternatively, the signal peptide can be linked to the protein of interest using a sequence which facilitates purification, such as with a GST domain.

In another embodiment, the signal peptides of the present invention can be used to identify regulatory sequences, *e.g.*, promoters, enhancers, repressors. Since signal peptides are the most amino-terminal sequences of a peptide, it is expected that the nucleic acids which flank the signal peptide on its amino-terminal side will be regulatory sequences which affect transcription. Thus, a nucleotide sequence which encodes all or a portion of a signal peptide can be used as a probe to identify and isolate signal peptides and their flanking regions, and these flanking regions can be studied to identify regulatory elements therein.

The present invention also pertains to variants of the polypeptides of the invention. Such variants have an altered amino acid sequence which can function as either agonists (mimetics) or as antagonists. Variants can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation. An agonist can retain substantially the same, or a subset, of the biological activities of the naturally

occurring form of the protein. An antagonist of a protein can inhibit one or more of the activities of the naturally occurring form of the protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the protein of interest. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein can have fewer side effects in a subject relative to treatment with the naturally occurring form of the protein.

Modification of the structure of the subject polypeptides can be for such purposes as enhancing therapeutic or prophylactic efficacy, stability (*e.g.*, *ex vivo* shelf life and resistance to proteolytic degradation *in vivo*), or post-translational modifications (*e.g.*, to alter phosphorylation pattern of protein). Such modified peptides, when designed to retain at least one activity of the naturally-occurring form of the protein, or to produce specific antagonists thereof, are considered functional equivalents of the polypeptides described in more detail herein. Such modified peptides can be produced, for instance, by amino acid substitution, deletion, or addition.

For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (*i.e.* isosteric and/or isoelectric mutations) will not have a major effect on the biological activity of the resulting molecule.

Whether a change in the amino acid sequence of a peptide results in a functional homolog (*e.g.*, functional in the sense that the resulting polypeptide mimics or antagonizes the wild-type form) can be readily determined by assessing the ability of the variant peptide to produce a response in cells in a fashion similar to the wild-type protein, or competitively inhibit such a response. Polypeptides in which more than one replacement has taken place can readily be tested in the same manner.

Variants of a protein of the invention which function as either agonists (mimetics) or as antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the protein of the invention for agonist or

antagonist activity. In one embodiment, a variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences
5 such that a degenerate set of potential protein sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display). There are a variety of methods which can be used to produce libraries of potential variants of the polypeptides of the invention from a degenerate oligonucleotide sequence. Methods for synthesizing degenerate oligonucleotides are
10 known in the art (*see, e.g.*, Narang, 1983, *Tetrahedron* 39:3; Itakura et al., 1984, *Annu. Rev. Biochem.* 53:323; Itakura et al., 1984, *Science* 198:1056; Ike et al., 1983, *Nucleic Acid Res.* 11:477).

In addition, libraries of fragments of the coding sequence of a polypeptide of the invention can be used to generate a variegated population of polypeptides for
15 screening and subsequent selection of variants. For example, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of the coding sequence of interest with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from
20 different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the protein of interest.

Several techniques are known in the art for screening gene products of
25 combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the
30 combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected.

Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify variants of a protein of the invention (Arkin and Yourvan, 1992, *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave et al., 1993, *Protein Engineering* 5 6(3):327-331).

An isolated polypeptide of the invention, or a fragment thereof can be used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. The full-length polypeptide or protein can be used or, alternatively, the invention provides antigenic peptide fragments for use as 10 immunogens. The antigenic peptide of a protein of the invention comprises at least 8 (preferably 10, 15, 20, or 30) amino acid residues of the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, and encompasses an epitope of the protein such that an antibody raised against the peptide forms a specific immune complex with the protein.

15 Preferred epitopes encompassed by the antigenic peptide are regions that are located on the surface of the protein, *e.g.*, hydrophilic regions. Hydropathy plots or similar analyses can be used to identify hydrophilic regions.

An isolated polypeptide of the invention, or a fragment thereof can be used as an immunogen to generate antibodies using standard techniques for polyclonal and 20 monoclonal antibody preparation. The full-length polypeptide or protein can be used or, alternatively, the invention provides antigenic peptide fragments for use as immunogens. The antigenic peptide of a protein of the invention comprises at least 8 (preferably 10, 15, 20, or 30) amino acid residues of the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, and encompasses an epitope of the protein 25 such that an antibody raised against the peptide forms a specific immune complex with the protein.

An immunogen typically is used to prepare antibodies by immunizing a suitable subject, (*e.g.*, rabbit, goat, mouse or other mammal). An appropriate immunogenic preparation can contain, for example, recombinantly expressed or 30 chemically synthesized polypeptide. The preparation can further include an adjuvant,

such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent.

Accordingly, another aspect of the invention pertains to antibodies directed against a polypeptide of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds an antigen, such as a polypeptide of the invention, *e.g.*, an epitope of a polypeptide of the invention. A molecule which specifically binds to a given polypeptide of the invention is a molecule which binds the polypeptide, but does not substantially bind other molecules in a sample, *e.g.*, a biological sample, which naturally contains the polypeptide. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')₂ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a polypeptide of the invention as an immunogen. Preferred polyclonal antibody compositions are ones that have been selected for antibodies directed against a polypeptide or polypeptides of the invention. Particularly preferred polyclonal antibody preparations are ones that contain only antibodies directed against a polypeptide or polypeptides of the invention. Particularly preferred immunogen compositions are those that contain no other human proteins such as, for example, immunogen compositions made using a non-human host cell for recombinant expression of a polypeptide of the invention. In such a manner, the only human epitope or epitopes recognized by the resulting antibody compositions raised against this immunogen will be present as part of a polypeptide or polypeptides of the invention.

The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA)

using immobilized polypeptide. If desired, the antibody molecules can be isolated from the mammal (*e.g.*, from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction.

Alternatively, antibodies specific for a protein or polypeptide of the invention can be
5 selected for (*e.g.*, partially purified) or purified by, *e.g.*, affinity chromatography. For example, a recombinantly expressed and purified (or partially purified) protein of the invention is produced as described herein, and covalently or non-covalently coupled to a solid support such as, for example, a chromatography column. The column can then be used to affinity purify antibodies specific for the proteins of the invention
10 from a sample containing antibodies directed against a large number of different epitopes, thereby generating a substantially purified antibody composition, *i.e.*, one that is substantially free of contaminating antibodies. By a substantially purified antibody composition is meant, in this context, that the antibody sample contains at most only 30% (by dry weight) of contaminating antibodies directed against epitopes
15 other than those on the desired protein or polypeptide of the invention, and preferably at most 20%, yet more preferably at most 10%, and most preferably at most 5% (by dry weight) of the sample is contaminating antibodies. A purified antibody composition means that at least 99% of the antibodies in the composition are directed against the desired protein or polypeptide of the invention.

20 At an appropriate time after immunization, *e.g.*, when the specific antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein, 1975, *Nature* 256:495-497, the human B cell hybridoma technique (Kozbor et al., 1983, *Immunol. Today* 4:72), the
25 EBV-hybridoma technique (Cole et al., 1985, *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pgs. 77-96) or trioma techniques. The technology for producing hybridomas is well known (*see generally Current Protocols in Immunology*, 1994, Coligan et al. (eds.) John Wiley & Sons, Inc., New York, NY). Hybridoma cells producing a monoclonal antibody of the invention are detected by
30 screening the hybridoma culture supernatants for antibodies that bind the polypeptide of interest, *e.g.*, using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody directed against a polypeptide of the invention can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with the polypeptide of interest. Kits for
5 generating and screening phage display libraries are commercially available (e.g., the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAPJ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No.
10 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al., 1991, *Bio/Technology* 9:1370-1372; Hay et al., 1992, *Hum. Antibod. Hybridomas* 3:81-85;
15 Huse et al., 1989, *Science* 246:1275-1281; Griffiths et al., 1993, *EMBO J.* 12:725-734.

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the
20 invention. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. (See, e.g., Cabilly et al., U.S. Patent No. 4,816,567; and Boss et al., U.S. Patent No. 4,816,397, which are incorporated herein by reference in their entirety). Such chimeric and humanized
25 monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT Publication No. WO 87/02671; European Patent Application 184,187; European Patent Application 171,496; European Patent Application 173,494; PCT Publication No. WO 86/01533; U.S. Patent No. 4,816,567; European Patent Application 125,023; Better et al., 1988,
30 *Science* 240:1041-1043; Liu et al., 1987, *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu et al., 1987, *J. Immunol.* 139:3521-3526; Sun et al., 1987, *Proc. Natl. Acad. Sci.*

USA 84:214-218; Nishimura et al., 1987, *Canc. Res.* 47:999-1005; Wood et al., 1985, *Nature* 314:446-449; and Shaw et al., 1988, *J. Natl. Cancer Inst.* 80:1553-1559); Morrison, 1985, *Science* 229:1202-1207; Oi et al., 1986, *Bio/Techniques* 4:214; U.S. Patent 5,225,539; Jones et al., 1986, *Nature* 321:552-525; Verhoeyan et al., 1988, 5 *Science* 239:1534; and Beidler et al., 1988, *J. Immunol.* 141:4053-4060.

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced, for example, using transgenic mice which are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain 10 genes. The transgenic mice are immunized in the normal fashion with a selected antigen, *e.g.*, all or a portion of a polypeptide of the invention. Monoclonal antibodies directed against the antigen can be obtained using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and 15 somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, *Int. Rev. Immunol.* 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such 20 antibodies, *see, e.g.*, U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806. In addition, companies such as Abgenix, Inc. (Freemont, CA), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

Completely human antibodies which recognize a selected epitope can be 25 generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, *e.g.*, a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope. (Jespers et al., 1994, *Bio/technology* 12:899-903).

An antibody directed against a polypeptide of the invention (*e.g.*, monoclonal 30 antibody) can be used to isolate the polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, such an antibody can be

used to detect the protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. The antibodies can also be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment
5 regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, 8-galactosidase, or acetylcholinesterase; examples of suitable
10 prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of
15 suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Further, an antibody (or fragment thereof) can be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin,
20 etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine,
25 cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly
30 actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a
5 toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, α -interferon, β -interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-
10 CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

Techniques for conjugating such therapeutic moiety to antibodies are well known, see, e.g., Arnon et al., "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in Monoclonal Antibodies And Cancer Therapy, Reisfeld et al. (eds.), pgs. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom et al., "Antibodies For
15 Drug Delivery", in Controlled Drug Delivery (2nd Ed.), Robinson et al. (eds.), pgs. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in Monoclonal Antibodies '84: Biological And Clinical Applications, Pinchera et al. (eds.), pgs. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer
20 Therapy", in Monoclonal Antibodies For Cancer Detection And Therapy, Baldwin et al. (eds.), pgs. 303-16 (Academic Press 1985), and Thorpe et al., 1982, "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", *Immunol. Rev.*, 62:119-58.

Alternatively, an antibody can be conjugated to a second antibody to form an
25 antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

Accordingly, in one aspect, the invention provides substantially purified antibodies or fragment thereof, and human and non-human antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid
30 sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, or an amino acid sequence encoded by the cDNA of a clone deposited as ATCC® Accession

Number 207220, 207223, PTA-248 or PTA-293; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, an amino acid sequence which is at least 95% identical to the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, wherein the
5 percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 or the BESTFIT program with BLOSUM 62 scoring matrix, gap open penalty of 12, frame shift penalty of 5, gap extend penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic
10 acid molecule consisting of any one of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, or the cDNA of a clone deposited as ATCC® Accession Number 207220, 207223, PTA-248 or PTA-293, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. In various embodiments, the substantially purified antibodies of the invention, or
15 fragments thereof, can be human, non-human, chimeric and/or humanized antibodies.

In another aspect, the invention provides human or non-human antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, or an
20 amino acid sequence encoded by the cDNA of a clone deposited as ATCC® Accession Number 207220, 207223, PTA-248 or PTA-293; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, an amino acid sequence which is at least 95% identical to the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, wherein the
25 percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 or the BESTFIT program with a BLOSUM 62 scoring matrix, gap open penalty of 12, frame shift penalty of 5, gap extend penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic
30 acid molecule consisting of any one of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16,

18, 19, 22, 24, or 25, or the cDNA of a clone deposited as ATCC® Accession Number 207220, 207223, PTA-248 or PTA-293, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C.

Such non-human antibodies can be goat, mouse, sheep, horse, chicken, rabbit, or rat
5 antibodies. Alternatively, the non-human antibodies of the invention can be chimeric and/or humanized antibodies. In addition, the non-human antibodies of the invention can be polyclonal antibodies or monoclonal antibodies.

In still a further aspect, the invention provides monoclonal antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide
10 comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, or an amino acid sequence encoded by the cDNA of a clone deposited as ATCC® Accession Number 207220, 207223, PTA-248 or PTA-293; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20,
15 23, or 91, an amino acid sequence which is at least 95% identical to the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 or the BESTFIT program with a BLOSUM 62 scoring matrix, gap open
20 penalty of 12, frame shift penalty of 5, gap extend penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of any one of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, or the cDNA of a clone deposited as any of ATCC® Accession Number 207220, 207223, PTA-248 or PTA-293, or a complement thereof, under
25 conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. The monoclonal antibodies can be human, humanized, chimeric and/or non-human antibodies.

The substantially purified antibodies or fragments thereof specifically bind to a signal peptide, a secreted sequence, an extracellular domain, a transmembrane or a
30 cytoplasmic domain cytoplasmic membrane of a polypeptide of the invention. In a

particularly preferred embodiment, the substantially purified antibodies or fragments thereof, the non-human antibodies or fragments thereof, and/or the monoclonal antibodies or fragments thereof of the invention specifically bind to a secreted sequence or an extracellular domain of the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91. Preferably, the secreted sequence or extracellular domain to which the antibody, or fragment thereof binds comprises from about amino acids 27 to 119 of SEQ ID NO:2, amino acid residues 1 to 305 of SEQ ID NO:5, amino acid residues 21 to 304 of SEQ ID NO:14, amino acid residues 20 to 318 of SEQ ID NO:17, or amino acid residues 17 to 141 of SEQ ID NO:20.

10 Any of the antibodies of the invention can be conjugated to a therapeutic moiety or to a detectable substance. Non-limiting examples of detectable substances that can be conjugated to the antibodies of the invention are an enzyme, a prosthetic group, a fluorescent material, a luminescent material, a bioluminescent material, and a radioactive material.

15 The invention also provides a kit containing an antibody of the invention conjugated to a detectable substance, and instructions for use. Still another aspect of the invention is a pharmaceutical composition comprising an antibody of the invention and a pharmaceutically acceptable carrier. In preferred embodiments, the pharmaceutical composition contains an antibody of the invention, a therapeutic moiety, and a pharmaceutically acceptable carrier.

Still another aspect of the invention is a method of making an antibody that specifically recognizes TANGO 244, TANGO 246, TANGO 275, TANGO 300, and MANGO 245, the method comprising immunizing a mammal with a polypeptide. The polypeptide used as an immunogen comprises an amino acid sequence selected from the group consisting of: the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, or an amino acid sequence encoded by the cDNA of a clone deposited as ATCC® Accession Number 207220, 207223, PTA-248 or PTA-293; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, an amino acid sequence which is at least 95% identical to the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, wherein the percent identity is determined using the ALIGN

- program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 or the BESTFIT program with a BLOSUM 62 scoring matrix, gap open penalty of 12, frame shift penalty of 5, gap extend penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of any one of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, or 25, or the cDNA of a clone deposited as ATCC® Accession Number 207220, 207223, PTA-248 or PTA-293, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. After immunization, a sample is collected from the mammal that contains an antibody that specifically recognizes GPVI. Preferably, the polypeptide is recombinantly produced using a non-human host cell. Optionally, the antibodies can be further purified from the sample using techniques well known to those of skill in the art. The method can further comprise producing a monoclonal antibody-producing cell from the cells of the mammal.
- Optionally, antibodies are collected from the antibody-producing cell.

III. Recombinant Expression Vectors and Host Cells

- Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide of the invention (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, expression vectors, are capable of directing the expression of genes to which they are

operably linked. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids (vectors). However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which

5 serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is

10 operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term

15 "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence

20 in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be

25 introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein.

The recombinant expression vectors of the invention can be designed for expression of a polypeptide of the invention in prokaryotic (*e.g.*, *E. coli*) or eukaryotic cells (*e.g.*, insect cells (using baculovirus expression vectors), yeast cells or

30 mammalian cells). Suitable host cells are discussed further in Goeddel, *supra*.

Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of
5 either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity
10 purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression
15 vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988, *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include
20 pTrc (Amann et al., 1988, *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion
25 promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to
30 express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, *Gene Expression Technology: Methods in*

Enzymology 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al., 1992, *Nucleic Acids Res.* 20:2111-2118).

- 5 Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari et al., 1987, *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz, 1982, *Cell* 30:933-10 943), pJRY88 (Schultz et al., 1987, *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and pPicZ (Invitrogen Corp, San Diego, CA).

Alternatively, the expression vector is a baculovirus expression vector. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al., 1983, *Mol. Cell Biol.* 3:2156-2165) 15 and the pVL series (Lucklow and Summers, 1989, *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987, *Nature* 329:840) and pMT2PC (Kaufman et al., 1987, *EMBO J.* 6:187-195). When used in mammalian cells, the 20 expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook et al., *supra*.

In another embodiment, the recombinant mammalian expression vector is 25 capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al., 1987, *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame 30 and Eaton, 1988, *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989, *EMBO J.* 8:729-733) and immunoglobulins

(Banerji et al., 1983, *Cell* 33:729-740; Queen and Baltimore, 1983, *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989, *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund et al., 1985, *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey
5 promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss, 1990, *Science* 249:374-379) and the α -fetoprotein promoter (Campes and Tilghman, 1989, *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a
10 DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operably linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to the mRNA encoding a polypeptide of the invention. Regulatory sequences operably linked to a nucleic acid cloned in the
15 antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in
20 which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub et al. (1986, *Reviews - Trends in Genetics*, Vol. 1).

Another aspect of the invention pertains to host cells into which a recombinant
25 expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may
30 not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic (*e.g.*, *E. coli*) or eukaryotic cell (*e.g.*, insect cells, yeast or mammalian cells).

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*supra*), and other laboratory manuals.

- 10 For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, for resistance to antibiotics) is generally introduced into the host cells along with the gene of interest.
- 15 Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

- In another embodiment, the expression characteristics of an endogenous (*e.g.*, TANGO 244, TANGO 246, TANGO 275, TANGO 300, and MANGO 245) nucleic acid within a cell, cell line or microorganism may be modified by inserting a DNA regulatory element heterologous to the endogenous gene of interest into the genome of a cell, stable cell line or cloned microorganism such that the inserted regulatory element is operatively linked with the endogenous gene (*e.g.*, TANGO 244, TANGO 246, TANGO 275, TANGO 300, and MANGO 245) and controls, modulates or activates the endogenous gene. For example, endogenous TANGO 244, TANGO 246, TANGO 275, TANGO 300, and MANGO 245 which are normally "transcriptionally silent", *i.e.*, TANGO 244, TANGO 246, TANGO 275, TANGO 300, and MANGO 245 genes which are normally not expressed, or are expressed only at very low levels in a cell line or microorganism, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally

expressed gene product in that cell line or microorganism. Alternatively, transcriptionally silent, endogenous TANGO 244, TANGO 246, TANGO 275, TANGO 300, and MANGO 245 genes may be activated by insertion of a promiscuous regulatory element that works across cell types.

- 5 A heterologous regulatory element may be inserted into a stable cell line or cloned microorganism, such that it is operatively linked with and activates expression of endogenous TANGO 244, TANGO 246, TANGO 275, TANGO 300, and MANGO 245 genes, using techniques, such as targeted homologous recombination, which are well known to those of skill in the art, and described *e.g.*, in Chappel, U.S. Patent No. 10 5,272,071; PCT publication No. WO 91/06667, published May 16, 1991.

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce a polypeptide of the invention. Accordingly, the invention further provides methods for producing a polypeptide of the invention using the host cells of the invention. In one embodiment, the method comprises culturing 15 the host cell of invention (into which a recombinant expression vector encoding a polypeptide of the invention has been introduced) in a suitable medium such that the polypeptide is produced. In another embodiment, the method further comprises isolating the polypeptide from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman 20 transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which a sequences encoding a polypeptide of the invention have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous sequences encoding a polypeptide of the invention have been introduced into their genome or homologous 25 recombinant animals in which endogenous encoding a polypeptide of the invention sequences have been altered. Such animals are useful for studying the function and/or activity of the polypeptide and for identifying and/or evaluating modulators of polypeptide activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one 30 or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians,

etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologous
5 recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

10 A transgenic animal of the invention can be created by introducing nucleic acid encoding a polypeptide of the invention (or a homologue thereof) into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Intronic sequences and polyadenylation signals can also be included in the transgene to
15 increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the polypeptide of the invention to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S.
20 Patent NOs. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of mRNA encoding the transgene in
25 tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying the transgene can further be bred to other transgenic animals carrying other transgenes.

To create an homologous recombinant animal, a vector is prepared which contains at least a portion of a gene encoding a polypeptide of the invention into
30 which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the gene. In a preferred embodiment, the vector is designed such

that, upon homologous recombination, the endogenous gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous gene is mutated or otherwise altered but still encodes functional

5 protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous protein). In the homologous recombination vector, the altered portion of the gene is flanked at its 5' and 3' ends by additional nucleic acid of the gene to allow for homologous recombination to occur between the exogenous gene carried by the vector and an endogenous gene in an embryonic stem cell. The

10 additional flanking nucleic acid sequences are of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see, e.g., Thomas and Capecchi, 1987, *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by

15 electroporation) and cells in which the introduced gene has homologously recombined with the endogenous gene are selected (see, e.g., Li et al., 1992, *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see, e.g., Bradley in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, ed. (IRL, Oxford, 1987) pgs. 113-152). A

20 chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and

25 homologous recombinant animals are described further in Bradley, 1991, *Current Opinion in Bio/Technology* 2:823-829 and in PCT Publication NOs. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One

30 example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso et al., 1992,

Proc. Natl. Acad. Sci. USA 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al., 1991, *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* 5 recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be 10 produced according to the methods described in Wilmut et al., 1997, *Nature* 385:810-813 and PCT Publication NOs. WO 97/07668 and WO 97/07669.

IV. Pharmaceutical Compositions

The nucleic acid molecules, polypeptides, and antibodies (also referred to 15 herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial 20 and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be 25 incorporated into the compositions.

The invention includes methods for preparing pharmaceutical compositions for modulating the expression or activity of a polypeptide or nucleic acid of the invention. Such methods comprise formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid 30 of the invention. Such compositions can further include additional active agents. Thus, the invention further includes methods for preparing a pharmaceutical

composition by formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid of the invention and one or more additional active compounds.

The agent which modulates expression or activity may, for example, be a
5 small molecule. For example, such small molecules include peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a
10 molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds. It is understood that appropriate doses of small molecule agents depends
15 upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small
20 molecule to have upon the nucleic acid or polypeptide of the invention. Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g. about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per
25 kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (e.g. a human) in order to modulate expression or activity of a polypeptide
30 or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose

until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor ELJ (BASF; Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures

thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

10 Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a polypeptide or antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

20 Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed.

25 Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal

silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser which contains a suitable
5 propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means.

For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile
10 salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with
15 conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems.
20 Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to
25 infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in
30 dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the

subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

For antibodies, the preferred dosage is 0.1 mg/kg to 100 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is described by Cruikshank et al. (1997, *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. Patent 5,328,470) or by stereotactic injection (see, e.g., Chen et al., 1994, *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) detection assays (*e.g.*, chromosomal mapping, tissue typing, forensic biology); c) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and d) methods of treatment (*e.g.*, therapeutic and prophylactic). For example, the TANGO 244, TANGO 246, TANGO 275, TANGO 300 and MANGO 245 polypeptides of the invention can be used to modulate cellular function, survival, morphology, proliferation, and/or differentiation of the cells in which they are expressed. The isolated nucleic acid molecules of the invention can be used to express proteins (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect mRNA (*e.g.*, in a biological sample) or a genetic lesion, and to modulate activity of a polypeptide of the invention. In addition, the polypeptides of the invention can be used to screen drugs or compounds which modulate activity or expression of a polypeptide of the invention as well as to treat disorders characterized by insufficient or excessive production of a protein of the invention or production of a form of a protein of the invention which has decreased or aberrant activity compared to the wild type protein. In addition, the antibodies of the invention can be used to detect and isolate a protein of the invention and modulate activity of a protein of the invention.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

A. Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) which bind to polypeptide of the invention or have a stimulatory or inhibitory effect on, for example, expression or activity of a polypeptide of the invention.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form

of a polypeptide of the invention or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; 5 synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, 1997, *Anticancer Drug Des.* 12:145).

- 10 Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al., 1993, *Proc. Natl. Acad. Sci. USA* 90:6909; Erb et al., 1994, *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann et al., 1994, *J. Med. Chem.* 37:2678; Cho et al., 1993, *Science* 261:1303; Carrell et al., 1994, *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell et al., 1994, *Angew. Chem. Int. Ed. Engl.* 33:2061; and Gallop et al., 1994, *J. Med. Chem.* 37:1233.

- Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992, *Bio/Techniques* 13:412-421), or on beads (Lam, 1991, *Nature* 354:82-84), chips (Fodor, 1993, *Nature* 364:555-556), bacteria (U.S. Patent No. 5,223,409), spores (Patent NOs. 5,571,698; 5,403,484; and 5,223,409), plasmids (Cull et al., 1992, *Proc.* 20 *Natl. Acad. Sci. USA* 89:1865-1869) or phage (Scott and Smith, 1990, *Science* 249:386-390; Devlin, 1990, *Science* 249:404-406; Cwirla et al., 1990, *Proc. Natl. Acad. Sci. USA* 87:6378-6382; and Felici, 1991, *J. Mol. Biol.* 222:301-310).

- In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of a polypeptide of the invention, or a biologically 25 active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to the polypeptide determined. The cell, for example, can be a yeast cell or a cell of mammalian origin. Determining the ability of the test compound to bind to the polypeptide can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding 30 of the test compound to the polypeptide or biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test

compounds can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the

5 enzymatic label detected by determination of conversion of an appropriate substrate to product. In a preferred embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test
10 compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially bind to the polypeptide or a biologically active portion thereof as compared to the known compound.

15 In another embodiment, the assay involves assessment of an activity characteristic of the polypeptide, wherein binding of the test compound with the polypeptide or a biologically active portion thereof alters (*e.g.*, increases or decreases) the activity of the polypeptide.

In another embodiment, an assay is a cell-based assay comprising contacting a
20 cell expressing a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of the polypeptide or a
25 biologically active portion thereof can be accomplished, for example, by determining the ability of the polypeptide protein to bind to or interact with a target molecule or to transport molecules across the cytoplasmic membrane.

Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can be accomplished by one of the methods described above
30 for determining direct binding. As used herein, a "target molecule" is a molecule with which a selected polypeptide (*e.g.*, a polypeptide of the invention binds or interacts

with in nature, for example, a molecule on the surface of a cell which expresses the selected protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A target molecule can be a polypeptide of the invention or some other polypeptide or protein. For example, a target molecule can be a component of a signal transduction pathway which facilitates transduction of an extracellular signal (*e.g.*, a signal generated by binding of a compound to a polypeptide of the invention) through the cell membrane and into the cell or a second intercellular protein which has catalytic activity or a protein which facilitates the association of downstream signaling molecules with a polypeptide of the invention. Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*e.g.*, intracellular Ca^{2+} , diacylglycerol, IP3, etc.), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene (*e.g.*, a regulatory element that is responsive to a polypeptide of the invention operably linked to a nucleic acid encoding a detectable marker, *e.g.* luciferase), or detecting a cellular response, for example, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with a test compound and determining the ability of the test compound to bind to the polypeptide or biologically active portion thereof. Binding of the test compound to the polypeptide can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound

to preferentially bind to the polypeptide or biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with a test
5 compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of the polypeptide can be accomplished, for example, by determining the ability of the polypeptide to bind to a target molecule by one of the methods described above for
10 determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of the polypeptide can be accomplished by determining the ability of the polypeptide of the invention to further modulate the target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

15 In yet another embodiment, the cell-free assay comprises contacting a polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to
20 interact with the polypeptide comprises determining the ability of the polypeptide to preferentially bind to or modulate the activity of a target molecule.

The cell-free assays of the present invention are amenable to use of both a soluble form or the membrane-bound form of a polypeptide of the invention. In the case of cell-free assays comprising the membrane-bound form of the polypeptide, it
25 may be desirable to utilize a solubilizing agent such that the membrane-bound form of the polypeptide is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-octylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton X-100, Triton X-114, Thesit, Isotridecypoly(ethylene glycol ether)_n, 3-[(3-
30 cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-

cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either the polypeptide of the invention or
5 its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to the polypeptide, or interaction of the polypeptide with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such
10 vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase fusion proteins or glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical; St. Louis, MO) or glutathione
15 derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or A polypeptide of the invention, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components
20 and complex formation is measured either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of binding or activity of the polypeptide of the invention can be determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the
25 screening assays of the invention. For example, either the polypeptide of the invention or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated polypeptide of the invention or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (*e.g.*, biotinylation kit, Pierce Chemicals; Rockford, IL), and immobilized in
30 the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with the polypeptide of the invention or target molecules but

which do not interfere with binding of the polypeptide of the invention to its target molecule can be derivatized to the wells of the plate, and unbound target or polypeptide of the invention trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-

- 5 immobilized complexes, include immunodetection of complexes using antibodies reactive with the polypeptide of the invention or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the polypeptide of the invention or target molecule.

- In another embodiment, modulators of expression of a polypeptide of the invention are identified in a method in which a cell is contacted with a candidate compound and the expression of the selected mRNA or protein (i.e., the mRNA or protein corresponding to a polypeptide or nucleic acid of the invention) in the cell is determined. The level of expression of the selected mRNA or protein in the presence of the candidate compound is compared to the level of expression of the selected
- 15 mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of expression of the polypeptide of the invention based on this comparison. For example, when expression of the selected mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified
- 20 as a stimulator of the selected mRNA or protein expression. Alternatively, when expression of the selected mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of the selected mRNA or protein expression. The level of the selected mRNA or protein expression in the cells can be determined by methods
- 25 described herein.

- In yet another aspect of the invention, a polypeptide of the inventions can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos et al., 1993, *Cell* 72:223-232; Madura et al., 1993, *J. Biol. Chem.* 268:12046-12054; Bartel et al., 1993, *Bio/Techniques* 14:920-924;
- 30 Iwabuchi et al., 1993, *Oncogene* 8:1693-1696; and PCT Publication No. WO 94/10300), to identify other proteins, which bind to or interact with the polypeptide of

the invention and modulate activity of the polypeptide of the invention. Such binding proteins are also likely to be involved in the propagation of signals by the polypeptide of the inventions as, for example, upstream or downstream elements of a signaling pathway involving the polypeptide of the invention.

- 5 This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

B. Detection Assays

- Portions or fragments of the cDNA sequences identified herein (and the
10 corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These
15 applications are described in the subsections below.

1. Chromosome Mapping

- Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome.
20 Accordingly, nucleic acid molecules described herein or fragments thereof can be used to map the location of the corresponding genes on a chromosome. The mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

- Briefly, genes can be mapped to chromosomes by preparing PCR primers
25 (preferably 15 to 25 bp in length) from the sequence of a gene of the invention. Computer analysis of the sequence of a gene of the invention can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only
30 those hybrids containing the human gene corresponding to the gene sequences will

yield an amplified fragment. For a review of this technique, see D'Eustachio et al. (1983, *Science* 220:919-924).

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the nucleic acid sequences of the invention to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a gene to its chromosome include *in situ* hybridization (described in Fan et al., 1990, *Proc. Natl. Acad. Sci. USA* 87:6223-27), pre-screening with labeled flow-sorted chromosomes (CITE), and pre-selection by hybridization to chromosome specific cDNA libraries. Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. For a review of this technique, see Verma et al., (Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York, 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, e.g., Egeland et al., 1987, *Nature* 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with a gene of the invention can be determined. If a mutation is observed in some or all of the affected individuals but not in any

unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes such as deletions or translocations that are visible from chromosome spreads or detectable using PCR
5 based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Furthermore, the nucleic acid sequences disclosed herein can be used to perform searches against "mapping databases", *e.g.*, BLAST-type search, such that the
10 chromosome position of the gene is identified by sequence homology or identity with known sequence fragments which have been mapped to chromosomes.

A polypeptide and fragments and sequences thereof and antibodies specific thereto can be used to map the location of the gene encoding the polypeptide on a chromosome. This mapping can be carried out by specifically detecting the presence
15 of the polypeptide in members of a panel of somatic cell hybrids between cells of a first species of animal from which the protein originates and cells from a second species of animal and then determining which somatic cell hybrid(s) expresses the polypeptide and noting the chromosome(s) from the first species of animal that it contains. For examples of this technique, see Pajunen et al., 1988, *Cytogenet. Cell*
20 *Genet.* 47:37-41 and Van Keuren et al., 1986, *Hum. Genet.* 74:34-40. Alternatively, the presence of the polypeptide in the somatic cell hybrids can be determined by assaying an activity or property of the polypeptide, for example, enzymatic activity, as described in Bordelon-Riser et al., 1979, *Somatic Cell Genetics* 5:597-613 and Owerbach et al., 1978, *Proc. Natl. Acad. Sci. USA* 75:5640-5644.

25

2. Tissue Typing

The nucleic acid sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphisms (RFLP)
30 for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield

unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

5 Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the nucleic acid sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and
10 subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals
15 and from tissue. The nucleic acid sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some
20 degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NOs:1, 4, 7, 10, 13, 16, 19, 22 and 25 can comfortably provide positive individual identification with a panel of perhaps 10 to
25 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NOs:3, 6, 9, 12, 15, 18, 21 and 24 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from the nucleic acid sequences described herein is used
30 to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification

database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial Gene Sequences in Forensic Biology

5 DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues,
10 *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can
15 enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions are
20 particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the nucleic acid sequences of the invention or portions thereof, *e.g.*, fragments derived from noncoding regions having a length of at least 20 or 30 bases.

25 The nucleic acid sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such probes can be used to identify tissue by
30 species and/or by organ type.

C. Predictive Medicine:

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically.

- 5 Accordingly, one aspect of the present invention relates to diagnostic assays for determining TANGO 244, TANGO 246, TANGO 275, TANGO 300, or MANGO 245 protein and/or nucleic acid expression as well as TANGO 244, TANGO 246, TANGO 275, TANGO 300, or MANGO 245 activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is
- 10 afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant or unwanted TANGO 244, TANGO 246, TANGO 275, TANGO 300, or MANGO 245 expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with TANGO 244, TANGO 246, TANGO 275, TANGO 300, or
- 15 MANGO 245 protein, nucleic acid expression or activity. For example, mutations in a TANGO 244, TANGO 246, TANGO 275, TANGO 300, or MANGO 245 gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with TANGO 244, TANGO 246, TANGO
- 20 275, TANGO 300, or MANGO 245 protein, nucleic acid expression or activity.

- As an alternative to making determinations based on the absolute expression level of selected genes, determinations may be based on the normalized expression levels of these genes. Expression levels are normalized by correcting the absolute expression level of a TANGO 244, TANGO 246, TANGO 275, TANGO
- 25 300, or MANGO 245 gene by comparing its expression to the expression of a gene that is not a TANGO 244, TANGO 246, TANGO 275, TANGO 300, or MANGO 245 gene, *e.g.*, a housekeeping gene that is constitutively expressed. Suitable genes for normalization include housekeeping genes such as the actin gene. This normalization allows the comparison of the expression level in one sample, *e.g.*, a patient sample, to
- 30 another sample, *e.g.*, a non-disease sample, or between samples from different sources.

Alternatively, the expression level can be provided as a relative expression level. To determine a relative expression level of a gene, the level of expression of the gene is determined for 10 or more samples of different cell isolates, preferably 50 or more samples, prior to the determination of the expression level for the sample in question. The mean expression level of each of the genes assayed in the larger number of samples is determined and this is used as a baseline expression level for the gene(s) in question. The expression level of the gene determined for the test sample (absolute level of expression) is then divided by the mean expression value obtained for that gene. This provides a relative expression level and aids in identifying extreme cases of disease.

Preferably, the samples used in the baseline determination will be from diseased or from non-diseased cells. The choice of the cell source is dependent on the use of the relative expression level. Using expression found in normal tissues as a mean expression score aids in validating whether the TANGO 244, TANGO 246, TANGO 275, TANGO 300, or MANGO 245 gene assayed is cell-type specific (versus normal cells). Such a use is particularly important in identifying whether a TANGO 244, TANGO 246, TANGO 275, TANGO 300, or MANGO 245 gene can serve as a target gene. In addition, as more data is accumulated, the mean expression value can be revised, providing improved relative expression values based on accumulated data. Expression data from cells provides a means for grading the severity of the disease state.

Another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of TANGO 244, TANGO 246, TANGO 275, TANGO 300, or MANGO 245 in clinical trials.

These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

An exemplary method for detecting the presence or absence of a polypeptide or nucleic acid of the invention in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting a polypeptide or nucleic acid (*e.g.*, mRNA, genomic

DNA) of the invention such that the presence of a polypeptide or nucleic acid of the invention is detected in the biological sample. A preferred agent for detecting mRNA or genomic DNA encoding a polypeptide of the invention is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic DNA encoding a polypeptide of the invention. The nucleic acid probe can be, for example, a full-length cDNA, such as the nucleic acid of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24 and 25, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to a mRNA or genomic DNA encoding a polypeptide of the invention.

10 Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting a polypeptide of the invention is an antibody capable of binding to a polypeptide of the invention, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of a polypeptide of the invention include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of genomic DNA include Southern hybridizations.

30 Furthermore, *in vivo* techniques for detection of a polypeptide of the invention include

introducing into a subject a labeled antibody directed against the polypeptide. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting a polypeptide of the invention or mRNA or genomic DNA encoding a polypeptide of the invention, such that the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide is detected in the biological sample, and comparing the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide in the control sample with the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide in the test sample.

The invention also encompasses kits for detecting the presence of a polypeptide or nucleic acid of the invention in a biological sample (a test sample). Such kits can be used to determine if a subject is suffering from or is at increased risk of developing a disorder associated with aberrant expression of a polypeptide of the invention (*e.g.*, a proliferative disorder, *e.g.*, psoriasis or cancer). For example, the kit can comprise a labeled compound or agent capable of detecting the polypeptide or mRNA encoding the polypeptide in a biological sample and means for determining the amount of the polypeptide or mRNA in the sample (*e.g.*, an antibody which binds the polypeptide or an oligonucleotide probe which binds to DNA or mRNA encoding the polypeptide). Kits can also include instructions for observing that the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of the polypeptide if the amount of the polypeptide or mRNA encoding the polypeptide is above or below a normal level.

For antibody-based kits, the kit can comprise, for example: (1) a first antibody (*e.g.*, attached to a solid support) which binds to a polypeptide of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

5 For oligonucleotide-based kits, the kit can comprise, for example: (1) an oligonucleotide, *e.g.*, a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule encoding a polypeptide of the invention. The kit can also comprise, *e.g.*, a buffering agent, a preservative, or a protein
10 stabilizing agent. The kit can also comprise components necessary for detecting the detectable agent (*e.g.*, an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit is usually enclosed within an individual container and all of the various containers are within a single package along with
15 instructions for observing whether the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of the polypeptide.

2. Prognostic Assays

The methods described herein can furthermore be utilized as diagnostic or
20 prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with aberrant expression or activity of a polypeptide
25 of the invention. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing such a disease or disorder. Thus, the present invention provides a method in which a test sample is obtained from a subject and a polypeptide or nucleic acid (*e.g.*, mRNA, genomic DNA) of the invention is detected, wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject
30 having or at risk of developing a disease or disorder associated with aberrant expression or activity of the polypeptide. As used herein, a "test sample" refers to a

biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, such methods can be used to determine whether a subject can be effectively treated with a specific agent or class of agents (*e.g.*, agents of a type which decrease activity of the polypeptide). Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant expression or activity of a polypeptide of the invention in which a test sample is obtained and the polypeptide or nucleic acid encoding the polypeptide is detected (*e.g.*, wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant expression or activity of the polypeptide).

The methods of the invention can also be used to detect genetic lesions or mutations in a gene of the invention, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized aberrant expression or activity of a polypeptide of the invention. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion or mutation characterized by at least one of an alteration affecting the integrity of a gene encoding the polypeptide of the invention, or the mis-expression of the gene encoding the polypeptide of the invention. For example, such genetic lesions or mutations can be detected by ascertaining the existence of at least one of: 1) a deletion of one or more nucleotides from the gene; 2) an addition of one or more nucleotides to the gene; 3) a substitution of one or more nucleotides of the gene; 4) a chromosomal rearrangement of the gene; 5) an alteration in the level of a messenger RNA transcript of the gene; 6) an aberrant modification of the gene, such as of the methylation pattern of the genomic DNA; 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; 8) a non-wild type level of a the

protein encoded by the gene; 9) an allelic loss of the gene; and 10) an inappropriate post-translational modification of the protein encoded by the gene. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a gene.

5 In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent NOs. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran et al., 1988, *Science* 241:1077-1080; and Nakazawa et al., 1994, *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter
10 of which can be particularly useful for detecting point mutations in a gene (*see, e.g.*, Abravaya et al., 1995, *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to the selected gene
15 under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting
20 mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli et al., 1990, *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, et al., 1989, *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi et al., 1988, *Bio/Technology* 6:1197), or any other nucleic
25 acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a selected gene from a sample cell
30 can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more

restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*, U.S. Patent No. 5,498,531) can be used to
5 score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin et al., 1996,
10 *Human Mutation* 7:244-255; Kozal et al., 1996, *Nature Medicine* 2:753-759). For example, genetic mutations can be identified in two-dimensional arrays containing light-generated DNA probes as described in Cronin et al., *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear
15 arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the
20 other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the selected gene and detect mutations by comparing the sequence of the sample nucleic acids with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on
25 techniques developed by Maxim and Gilbert (1977, *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger (1977, *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (1995, *Bio/Techniques* 19:448), including sequencing by mass spectrometry (*see, e.g.*, PCT Publication No. WO 94/16101;
30 Cohen et al., 1996, *Adv. Chromatogr.* 36:127-162; and Griffin et al., 1993, *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in a selected gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al., 1985, *Science* 230:1242). In general, the technique of "mismatch cleavage" entails providing heteroduplexes
5 formed by hybridizing (labeled) RNA or DNA containing the wild-type sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. RNA/DNA duplexes can be treated with RNase to digest
10 mismatched regions, and DNA/DNA hybrids can be treated with S1 nuclease to digest mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to
15 determine the site of mutation. *See, e.g.*, Cotton et al., 1988, *Proc. Natl. Acad. Sci. USA* 85:4397; Saleeba et al., 1992, *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so
20 called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al., 1994, *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe
25 based on a selected sequence, *e.g.*, a wild-type sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.*, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to
30 identify mutations in genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant

and wild type nucleic acids (Orita et al., 1989, *Proc. Natl. Acad. Sci. USA* 86:2766; see also Cotton, 1993, *Mutat. Res.* 285:125-144; Hayashi, 1992, *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded
5 nucleic acids varies according to sequence, and the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the
10 subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al., 1991, *Trends Genet.* 7:5).

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing
15 gradient gel electrophoresis (DGGE) (Myers et al., 1985, *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a 'GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the
20 mobility of control and sample DNA (Rosenbaum and Reissner, 1987, *Biophys. Chem.* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in
25 which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al., 1986, *Nature* 324:163; Saiki et al., 1989, *Proc. Natl. Acad. Sci. USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the
30 hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; Gibbs et al., 1989, *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent or reduce polymerase extension (Prossner, 1993, *Tibtech* 11:238). In addition, it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al., 1992, *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany, 1991, *Proc. Natl. Acad. Sci. USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a gene encoding a polypeptide of the invention. Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which the polypeptide of the invention is expressed may be utilized in the prognostic assays described herein.

3. Pharmacogenomics

Agents, or modulators which have a stimulatory or inhibitory effect on activity or expression of a polypeptide of the invention as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders associated with aberrant activity of the polypeptide. In conjunction with such treatment, the pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by

altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of a polypeptide of the invention, expression of a nucleic acid of the invention, or mutation content of a gene of the invention in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. *See, e.g.*, Linder, 1997, *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body are referred to as "altered drug action." Genetic conditions transmitted as single factors altering the way the body acts on drugs are referred to as "altered drug metabolism". These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead

to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, a PM will show no therapeutic response, as demonstrated for the analgesic effect of codeine
5 mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of a polypeptide of the invention, expression of a nucleic
10 acid encoding the polypeptide, or mutation content of a gene encoding the polypeptide in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug
15 responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a modulator of activity or expression of the polypeptide, such as a modulator identified by one of the exemplary screening assays described herein.

20

4. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of a polypeptide of the invention (*e.g.*, the ability to modulate aberrant cell proliferation chemotaxis, and/or differentiation) can be applied not only in basic drug
25 screening, but also in clinical trials. For example, the effectiveness of an agent, as determined by a screening assay as described herein, to increase gene expression, protein levels or protein activity, can be monitored in clinical trials of subjects exhibiting decreased gene expression, protein levels, or protein activity.

Alternatively, the effectiveness of an agent, as determined by a screening assay, to
30 decrease gene expression, protein levels or protein activity, can be monitored in clinical trials of subjects exhibiting increased gene expression, protein levels, or

protein activity. In such clinical trials, expression or activity of a polypeptide of the invention and preferably, that of other polypeptide that have been implicated in for example, a cellular proliferation disorder, can be used as a marker of the immune responsiveness of a particular cell.

5 For example, and not by way of limitation, genes, including those of the invention, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) which modulates activity or expression of a polypeptide of the invention (*e.g.*, as identified in a screening assay described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for
10 example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of a gene of the invention and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described
15 herein, or by measuring the levels of activity of a gene of the invention or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

20 In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to
25 administration of the agent; (ii) detecting the level of the polypeptide or nucleic acid of the invention in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level the of the polypeptide or nucleic acid of the invention in the post-administration samples; (v) comparing the level of the polypeptide or nucleic acid of the invention in the pre-administration
30 sample with the level of the polypeptide or nucleic acid of the invention in the post-administration sample or samples; and (vi) altering the administration of the agent to

the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of the polypeptide to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of the polypeptide to lower levels than detected, i.e., to decrease the effectiveness of the agent.

C. Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant expression or activity of a polypeptide of the invention, *e.g.*, cardiac infection (*e.g.*, myocarditis or dilated cardiomyopathy), central nervous system infection (*e.g.*, non-specific febrile illness or meningoencephalitis), pancreatic infection (*e.g.*, acute pancreatitis), respiratory infection (pneumonia), gastrointestinal infection, type I diabetes, cancer, familia hypercholesterolemia, treat hemophilia B, Marfan syndrome, protein S deficiency, allergy, inflammation, and gastroduodenal ulcer. Moreover, the polypeptides of the invention can be used to modulate cellular function, survival, morphology, proliferation and/or differentiation.

20 1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant expression or activity of a polypeptide of the invention, by administering to the subject an agent which modulates expression or at least one activity of the polypeptide. Subjects at risk for a disease which is caused or contributed to by aberrant expression or activity of a polypeptide of the invention can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of aberrancy, for example, an agonist or antagonist agent can be used for treating the subject.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating expression or activity of a polypeptide of the invention for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of the polypeptide. An agent that modulates activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of the polypeptide, a peptide, a peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more of the biological activities of the polypeptide. Examples of such stimulatory agents include the active polypeptide of the invention and a nucleic acid molecule encoding the polypeptide of the invention that has been introduced into the cell. In another embodiment, the agent inhibits one or more of the biological activities of the polypeptide of the invention. Examples of such inhibitory agents include antisense nucleic acid molecules and antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a polypeptide of the invention. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) expression or activity. In another embodiment, the method involves administering a polypeptide of the invention or a nucleic acid molecule of the invention as therapy to compensate for reduced or aberrant expression or activity of the polypeptide.

Stimulation of activity is desirable in situations in which activity or expression is abnormally low or downregulated and/or in which increased activity is likely to have a beneficial effect. Conversely, inhibition of activity is desirable in situations in which activity or expression is abnormally high or upregulated and/or in which decreased activity is likely to have a beneficial effect.

The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the

5 following Claims.

International Application No: PCT/ /

MICROORGANISMS

Optional Sheet in connection with the microorganism referred to on pages __, lines ____ of the description *

A. IDENTIFICATION OF DEPOSIT *

Further deposits are identified on an additional sheet *

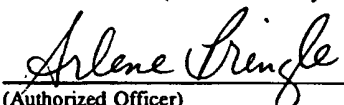
Name of depositary institution *

American Type Culture Collection

Address of depositary institution (including postal code and country) *

10801 University Blvd.
Manassas, VA 20110-2209
USDate of deposit * April 21, 1999 Accession Number * 207220**B. ADDITIONAL INDICATIONS *** (leave blank if not applicable). This information is continued on a separate attached sheet**C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *** (if the indications are not all designated States)**D. SEPARATE FURNISHING OF INDICATIONS *** (leave blank if not applicable)

The indications listed below will be submitted to the International Bureau later * (Specify the general nature of the indications e.g., "Accession Number of Deposit")

E. ☒ This sheet was received with the International application when filed (to be checked by the receiving Office)
(Authorized Officer)☐ The date of receipt (from the applicant) by the International Bureau *

was

(Authorized Officer)

Form PCT/RO/134 (January 1981)

International Application No: PCT/ /

Form PCT/RO/134 (cont.)

American Type Culture Collection

10801 University Blvd.
Manassas, VA 20110-2209
US

<u>Accession No.</u>	<u>Date of Deposit</u>
207223	April 21, 1999
207223	April 21, 1999
207223	April 21, 1999
PTA-248	June 18, 1999
PTA-293	June 30, 1999

What is claimed is:

1. An isolated nucleic acid molecule selected from the group consisting of:
 - 5 a) a nucleic acid molecule comprising a nucleotide sequence which is at least 55% identical to the nucleotide sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, 25, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207220, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the cDNA insert of the plasmid deposited
10 with the ATCC® as Accession Number PTA-248, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-293, or a complement thereof;
 - b) a nucleic acid molecule comprising a fragment of at least 300 nucleotides of the nucleotide sequence of SEQ ID NOs: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24 or 25, the cDNA insert of the plasmid deposited with the ATCC® as
15 Accession Number 207220, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-248, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-293, or a complement thereof;
 - c) a nucleic acid molecule which encodes a polypeptide comprising the
20 amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207220, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as
25 Accession Number PTA-248, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-293;
 - d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with
30 the ATCC® as Accession Number 207220, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223,

the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC[®] as Accession Number PTA-248, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC[®] as Accession Number PTA-293, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NOs: 2, 5, 8, 11, 14, 17, 20, 23, or 91, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC[®] as Accession Number 207220, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC[®] as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC[®] as Accession Number PTA-248, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC[®] as Accession Number PTA-293; and

e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NOs: 2, 5, 8, 11, 14, 17, 20, 23, or 91, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC[®] as Accession Number 207220, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC[®] as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC[®] as Accession Number PTA-248, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC[®] as Accession Number PTA-293, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NOs: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, 25, or a complement thereof, under stringent conditions.

2. The isolated nucleic acid molecule of Claim 1, which is selected from the group consisting of:

a) a nucleic acid comprising the nucleotide sequence of SEQ ID NOs: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24 or 25, the cDNA insert of the plasmid deposited with the ATCC[®] as Accession Number 207220, the cDNA insert of the plasmid deposited with the ATCC[®] as Accession Number 207223, the cDNA insert of the plasmid deposited with the ATCC[®] as Accession Number PTA-248, the cDNA

insert of the plasmid deposited with the ATCC® as Accession Number PTA-293, or a complement thereof; and

- b) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207220, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-248, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-293.

3. The nucleic acid molecule of Claim 1 further comprising vector nucleic acid sequences.

4. The nucleic acid molecule of Claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.

5. A host cell which contains the nucleic acid molecule of Claim 1.

6. The host cell of Claim 5 which is a mammalian host cell.

7. A non-human mammalian host cell containing the nucleic acid molecule of Claim 1.

8. An isolated polypeptide selected from the group consisting of:

- a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91;
- b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as

- Accession Number 207220, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-248, or the amino acid sequence encoded by the cDNA
- 5 insert of the plasmid deposited with the ATCC® as Accession Number PTA-293, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NOs: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, 25, or a complement thereof under stringent conditions; and
- c) a polypeptide which is encoded by a nucleic acid molecule comprising
- 10 a nucleotide sequence which is at least 55% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NOs: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, 25, or a complement thereof.
9. The isolated polypeptide of Claim 8 comprising the amino acid
- 15 sequence of SEQ ID NOs: 2, 5, 8, 11, 14, 17, 20, 23, or 91.
10. The polypeptide of Claim 8 further comprising heterologous amino acid sequences.
- 20 11. An antibody which selectively binds to a polypeptide of Claim 8.
12. A method for producing a polypeptide selected from the group consisting of:
- a) a polypeptide comprising the amino acid sequence of SEQ ID NOs: 2,
- 25 5, 8, 11, 14, 17, 20, 23, or 91, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207220, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-248, or the
- 30 amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-293;

b) a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207220, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-248, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-293, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207220, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-248, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-293; and

c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, or 91, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207220, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-248, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-293, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 22, 24, 25, or a complement thereof under stringent conditions;

comprising culturing the host cell of Claim 5 under conditions in which the nucleic acid molecule is expressed.

13. A method for detecting the presence of a polypeptide of Claim 8 in a sample, comprising:

- a) contacting the sample with a compound which selectively binds to a polypeptide of Claim 8; and
- 5 b) determining whether the compound binds to the polypeptide in the sample.

14. The method of Claim 13, wherein the compound which binds to the polypeptide is an antibody.

10

15. A kit comprising a compound which selectively binds to a polypeptide of Claim 8 and instructions for use.

16. A method for detecting the presence of a nucleic acid molecule of Claim 1 in a sample, comprising the steps of:

- 15 a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and
- b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample.

20

17. The method of Claim 16, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

18. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of Claim 1 and instructions for use.

25

19. A method for identifying a compound which binds to a polypeptide of Claim 8 comprising the steps of:

- a) contacting a polypeptide, or a cell expressing a polypeptide of Claim 8 with a test compound; and
- 30 b) determining whether the polypeptide binds to the test compound.

20. The method of Claim 19, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detecting of test compound/polypeptide binding;

5 b) detection of binding using a competition binding assay;

c) detection of binding using an assay for TANGO 244-, TANGO 246-, TANGO 275-, TANGO 300-, or MANGO 245-mediated signal transduction.

21. A method for modulating the activity of a polypeptide of Claim 8
10 comprising contacting a polypeptide or a cell expressing a polypeptide of Claim 8 with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

22. A method for identifying a compound which modulates the activity of
15 a polypeptide of Claim 8, comprising:

a) contacting a polypeptide of Claim 8 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

20

GTCCACCCACGGCTCCGGCCCCGACACCCAGACCGACCTTGACCGCCCCACCTGGCAGGAGCAGGACAGGACGGCCCGGACG 79
 M A E L P G P F L C G A L L G F L C
 CGGCC ATG GCC GAG CTC CCG GGG CCC TTT CTC TGC GGG GCC CTG CTA GGC TTC CTG TGC 138
 C S Y P P S N P L C S Q S G Q T S V G G
 CTG AGT GTT CCC CCC AGT AAT CCC TTA TGC AGT CAG AGT GGA CAA ACC TCT GTG GGA GGC 198
 S T A L R C S S S E G A P K P Y Y N W V
 TCT ACT GCA CTG AGA TGC AGC TCT TCC GAG GGG GCT CCT AAG CCA GTG TAC AAC TGG GTG 258
 R L G T F P T P S P G S M V Q D E V S G
 CGT CTT GGA ACT TTT CCT ACA CCT TCT CCT GGC AGC ATG GTT CAA GAT GAG GTG TCT GGC 318
 Q L I L T N L S L T S S G T Y R C V A T
 CAG CTC ATT CTC ACC AAC CTC TCC CTG ACC TCC TCG GGC ACC TAC CGC TGT GTG GCC ACC 378
 N Q M G S A S C E L T L S V T E P S Q G
 AAC CAG ATG GGC AGT GCA TCC TGT GAG CTG ACC CTC TCT GTG ACC GAA CCC TCC CAA GGC 438
 R Y A G A L I G Y L L G Y L L L S V A A
 CGA GTG GCC GGA GCT CTG ATT GGG GTG CTC CTG GGC GTG CTG TTG CTG TCA GTT GCT GCG 498
 F C L V R F Q K E R G K K P K E T Y G G
 TTC TGC CTG GTC AGG TTC CAG AAA GAG AGG GGG AAG AAG CCC AAG GAG ACA TAT GGG GGT 558
 S D L R
 AGT GAC CTT CGG TGA 573
 GCAGGAGGGCTGGGGGGTGGCGCAAGGAGGGAGGAAAGGGCTTGAGTTAAAAGCGGGTGCCTGCAACCCCTCAAACCTCCG 652
 ACATCATTCAAGTGTGTTTAGGGGCAGGAGGTGTTGTTTCAGCCGTGGAATTTGCTGGTGGCAGCAGTGTAACCTGTGTAT 731
 TTGAGGGTACAGGCAAGCGGTACAGGGTGGAGTGGCTGGTCCACAAGCTGTGGCAGGGAAGCTGTTTGCAGGACTGCC 810
 TGCCCCCTCCTCATATTTAATAAAAGTTTACTTTTCTGTTCCGAAGGTATTTTCATATATTTTAACCCCTGGGAGTAGTA 889
 GTGGCTTGTAGATGCCAGGAAATGGATTTGTCTGAGCAGTCAGCTGAGTTCAATTCTTCTGTGGAGGAAATCAGGAAA 968
 GGGGAGGGGAAACTGCCTCTGTTCATCCACTTTAGCTGCCAGNCAGGGTCTAGGATAGGGATCAGAGCAACATTTCTTCA 1047
 GGTGGAGTCTCAGATTACCTGGACAGAAATCACCGGGAAGTACTTATACATTTCAGATTTCAGGCCACTTCTAGCCTTCC 1126
 TGTAGTTGTGCGTTGGGGAGTGATNAGGCCCANAAATTTTCNTTTTAACCAAAGTTCCNCANATTATTTTCAAGCCAGT 1205
 GAAATTTAAGAGTCCCCAGGTTAGAGGACGGCCCTCCNCCGAGGAGGNTTTTACTGKTTACTCAGAACTTGCCTATAC 1284
 CCATCAGGGAGGATGCCATCGCTCCTGGGATCTCTGAGCACACTTGTATGAGGGCTGATTCTAGCAAGGGGTTCTCTGGA 1363
 AAGACCCTCGTCTGCCAGCACCGTGACGACCACCAAGTCCAAGCTCCCTATGGTCTGTGACTTCTCCCGATCCCTGAG 1442
 GGCGGTGAGGGGAATATCAATAATTAAAGTCTGTGGGTACCAAAAAAAAAAAAAAAAAAAGGGCGGCCGC 1513

FIGURE 1

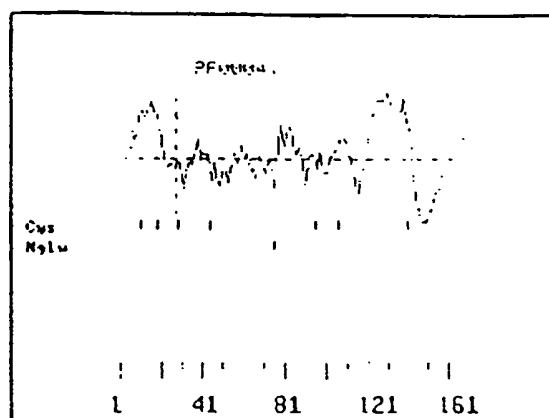


FIGURE 2

```

CONSENSUS      *-->GesvtLtCsvsqfgppgvsvtWyfkngk.lgpsilcysysriesgek
                G+s  L+Cs+s +g p + ++W  + g+ ++p      ++g
TANGO 244  37   GGSTALRCSSS-EGAPKPVYNWV-RLGTfPTP-----SPG-- 59

                anlsegrfsissitLtissvekeDsGtYtCvv<--
                +++ +++S++  E ++++++  sGtY+Cv+
70 -SMVQDEVSGQ---LILTNLSTSSGTYRCVA      97

```

FIGURE 3

ALIGN calculates a global alignment of two sequences
 version 2.0u Please cite: Myers and Miller, CABIOS (1989)
 > hT244 a.a. 162 aa vs.
 > GenPept AF061022 - Homo sapiens CTH gene, compl 125 aa
 scoring matrix: paml20.mat, gap penalties: -12/-4
 48.6% identity: Global alignment score: 149

```

      10      20
inputs MAELPGPFLLCGALLGFLCLS-----
      .....
      MADLPGPFLLCGALLGFLCLXLAVEVKVPTEPLSTPLGKTAELTCTYSTSVGDTFALEWSF
      10      20      30      40      50      60

inputs -----
      VQPGKPISESHPILYFTNGHLYPTGSKSKRVSLNQNPPTVG VATLKLTDVHPSDTGTYLC
      70      80      90      100     110     120

      30      40      50
inputs -----VPPSNPLCSQSGQTSVGGSTALRCSSEGAPKPVYNWVR
      .....
      QVNNPPDFYTNGLGLINLTVLVPPSNPLCSQSGQTSVGGSTALRCSSEGAPKPVYNWVR
      130     140     150     160     170     180

      60      70      80      90      100     110
inputs LGTFPTPSPGSMVQDEVSGQLILTNLSTSSGTYRCVATNQMGASCELTLSVTEPSQGR
      .....
      LGTFPTPSPGSMVQDEVSGQLILTNLSTSSGTYRCVATNQLGSASCELTLSVTEPSQGR
      190     200     210     220     230     240

      120     130     140     150     160
inputs VAGALIGVLLGVLLLSVAAFCLVRFQKERGKKPKETYGGSDLR-----
      .....
      VTGALIGVLLGVLLLSVAAFCLVRFQKERGKKPKETYGGSDLREDALAPGISSEHTCMRAD
      250     260     270     280     290     300

inputs -----
      SSKGFLERPSASTVTTTTSKSLPMVV
      310     320

```

FIGURE 4

Input file Athsa34d2.seq: Output File Athsa34d2.pat
Sequence length 2002

```

GTCCACCCACGCGTCCGCAACATCCTGGCTTAGTATTGTGTGCAAAATCAGAGAGGGGTGCAAGATCCTGATTTTTCAG 79
      M  A  A  Q  Y  G  S  M  S  F  N  P  S  T  P  G
GAGTTCAAGCGACA ATG GCA GCC CAA TAC GGC AGT ATG AGC TTC AAC CCC AGC ACA CCA GGG 141
  A  S  Y  G  P  G  R  Q  E  P  R  N  S  Q  L  R  I  V  L  V
GCC AGT TAT GGG CCT GGA AGG CAA GAG CCC AGA AAT TCC CAA TTG AGA ATT GTG TTA GTG 201
  G  K  T  G  A  G  K  S  A  T  G  N  S  I  L  G  R  K  V  F
GGT AAA ACC GGA GCA GGA AAA AGT GCA ACA GGA AAC AGC ATC CTT GGC CGG AAA GTG TTT 261
  H  S  G  T  A  A  K  S  I  T  K  K  C  E  K  R  S  S  S  W
CAT TCT GGC ACT GCA GCA AAA TCC ATT ACC AAG AAG TGT GAG AAA CGC AGC AGC TCA TGG 321
  K  E  T  E  L  V  V  V  D  T  P  G  I  F  D  T  E  V  P  N
AAG GAA ACA GAA CTT GTC GTA GTT GAC ACA CCA GGC ATT TTC GAC ACA GAG GTG CCC AAT 381
  A  E  T  S  K  E  I  I  R  C  I  L  L  T  S  P  G  P  H  A
GCT GAA ACG TCC AAG GAG ATT ATT CGC TGC ATT CTT CTG ACC TCC CCA GGG CCT CAT GCT 441
  L  L  L  V  V  P  L  G  R  Y  T  E  E  E  H  K  A  T  E  K
CTG CTT CTG GTG GTT CCA CTG GGC CGT TAC ACT GAG GAA GAG CAC AAA GCC ACA GAG AAG 501
  I  L  K  M  F  G  E  R  A  R  S  F  M  I  L  I  F  T  R  K
ATC CTG AAA ATG TTT GGA GAG AGG GCT AGA AGT TTC ATG ATT CTC ATA TTC ACC CGG AAA 561
  D  D  L  G  D  T  N  L  H  D  Y  L  R  E  A  P  E  D  I  Q
GAT GAC TTA GGT GAC ACC AAT TTG CAT GAC TAC TTA AGG GAA GCT CCA GAA GAC ATT CAA 621
  D  L  M  D  I  F  G  D  R  Y  C  A  L  N  N  K  A  T  G  A
GAC TTG ATG GAC ATT TTC GGT GAC CGC TAC TGT GCG TTA AAC AAC AAG GCA ACA GGC GCT 681
  E  Q  E  A  Q  R  A  Q  L  L  G  L  I  Q  R  V  V  R  E  N
GAG CAG GAG GCC CAG AGG GCA CAG TTG CTG GGC CTG ATC CAG CGC GTG GTG AGG GAG AAC 741
  K  E  G  C  Y  T  N  R  M  Y  Q  R  A  E  E  E  I  Q  K  Q
AAG GAA GGC TGC TAC ACT AAT AGG ATG TAC CAA AGG GCG GAG GAG GAG ATC CAG AAG CAA 801
  T  Q  A  M  Q  E  L  H  R  V  E  L  E  R  E  K  A  R  I  R
ACA CAA GCA ATG CAA GAA CTC CAC AGA GTG GAG CTG GAG AGA GAG AAA GCG CGG ATA AGA 861
  E  E  Y  E  E  K  I  R  K  L  E  D  K  V  E  Q  E  K  R  K
GAG GAG TAT GAA GAG AAA ATC AGA AAG CTG GAA GAT AAA GTG GAG CAG GAA AAG AGA AAG 921
  K  Q  M  E  K  K  L  A  E  Q  E  A  H  Y  A  V  R  Q  Q  R
AAG CAA ATG GAG AAG AAA CTA GCA GAA CAG GAG GCT CAC TAT GCT GTA AGG CAG CAA AGG 981
  A  R  T  E  V  E  S  K  D  G  I  L  E  L  I  M  T  A  L  Q
GCA AGA ACG GAA GTG GAG AGT AAG GAT GGG ATA CTT GAA TTA ATC ATG ACA GCG TTA CAG 1041
  I  A  S  F  I  L  L  R  L  F  A  E  D  *
ATT GCT TCC TTT ATT TTG TTA CGT CTG TTC GCG GAA GAT TAA 1083
ACTTAATGAAAATCTGTTTGTATTTTCTGCATATTCTCTGGCAACCTTGCCCCATACTTACTTATTTAGCATAGTCGAG 1162
TGCTCTAGTTTCTGTCTCTCAGGCACTCGTAACTAAGGACCACCATTTGGCCATTGGTAGATGTTTGATTGACTTAACAA 1241
GAGAGGGACAAATTTTCAATTTGTGAAACTCCAAAGCAGAAAGTATTGGTGCTTGCTACCTTGTGAATTCTTCCTTAGA 1320

```

FIGURE 5A

CATGCAGAGAAAATGTATGCAAGAGACCAAAAAGATGGCTCCAAGCTATGTCATGTTACCTGTAATAAAATCTTTTCTT 1399
CTAGATTCTTTCTATGTTGGCAGATAATCTCCCCCTGTAGCTTCCACTCACTTATTCTTGCATTCAGAGTCACAAATGAT 1478
CATCTTACCCATGTGGTTTTTTGAGAAAGAAAGATCAATTCTTTGTTTTGCAGTAGGTAATCTTAGAGATGGAGATGATTG 1557
TAGAATTATTCTTAGATGAGTGTCAATTTATTTAATTCCATTGTCATATAAGGAGTCAAATTGTTTCTTATCATTGTGTT 1636
CATTGAAGAACAGAGACCTGTCTGGAAAATCGATCTCTACAAATTCAATTAAATAATGATCCCCAAATGCTGAAAAAGT 1715
GAAATACAGCAATTCAACAGATAATAGAGCAATGTTTAGTATATTAGCTGTATCTGTAGAACTCTTTGACGAACCTC 1794
AATTTAACCAATTTGATGAATACCCAGTTCTCTCTTTTCTAGAGAAAGATAGTTGCAACCTCACCTCCCTCACTCAAC 1873
ACTTTGAATACTTATTGTTTGGCAGGTCATCCACACACTTCTGCCCCCACTGCATTGAATTTTTTGCTTATGTTGTTTA 1952
TAATAAACTTTTCAATTATCTCAAAAAAAAAAAAAAAAAAAAA 1992

FIGURE 5B

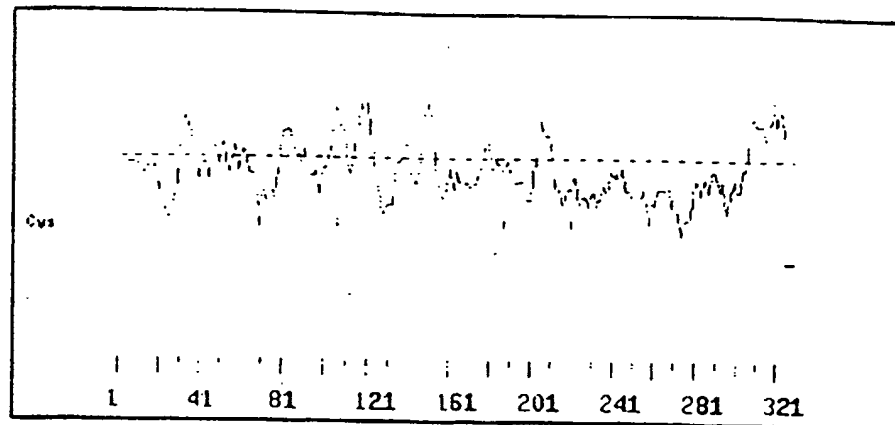


FIGURE 6

```

CONSENSUS      *->KGfdFTLMVVGesGLGKtTlINTLFLtdLidangvanDsreidgase
                +  ++ + +VG  G GK++  N+ + +++  + +          a++
TANGO 246      17  RNSQLRIVLVGKTGAGKSATGNSILGRKVFHSGT-----AAK 63

                tkikktveIkeitkveiEEdGvkLnLTViDTPG.FGDaiDNskcwepIve
                +i  ++ + ++ + +E          L V+DTPG F  - N+++ + I++
154 -SI--TKKCE-KRSSWKET----ELVVVDTPGiFDTEVPNAETSKEIIR 105

                YIdeQheqYLrqEsrinRtkivDnRVHcCLYFIspTghgLkpLDvefMKk
                                C+  sp  h L+ L v  + +
106 -----CILLTSPGPHALL-L-VVPLGR 125

                LsekVNIIPVIAKADtLTadElqefKkrIreei....erqnlkIYkFPde
                                +T +E + + ++I+  +++++ + + I I++
126 -----YTEEEHKAT-EKILKMFgeraRSFMILIFT--RK 156

                eeDeGDEefkeqtqqLkssiPFAIVGSneeIengdGekVRGRkYPWGvVE
                +++ D      +L++          + e+i++          + +
157 DDLG-DTN---LHDYLR-----APEDIQD-----LMD 180

                VENpsHCDFvkLRnlLirCHLqDLketTeeilyEnYRsekLsalglkaen
                +  +++C  +L n          k t +e+  E R++ L  ++ ++
181 IFGDRYC---ALNN-----KATGAEO--EAORAQLLGLIQRVVRE 215

```

FIGURE 7

```

CONSENSUS      *->GevialvGpNGaGKSTLLklisGllp.....pteGtilldGardlr
                +++lvG +GaGKS  + i+G  +++++  + t + + r+ +
TANGO 246  30   QLRIVLVGKTGAGKSATGNSILGRKVfhsgtaAKSITKKCEK-RSSS 75

                .lsklkerlerlrknigvvfQdptlfpnveltvreniafgirls.....
                +      e+      + + + ++t  pn  t +e+i+  l  s++++
76  wKE-----TELVV--VDTPGIFDTEVPN-AETSKEIIRCILLTSpgphal 117

                .....lglskdeqrarlkkgaeelLerlglydhldrrpgtLSGGqk
                +      +++++e +a      e++L+++g      +++
118  llvvplGRYTEEEHKAT-----EKILKMFGER-----ARS----- 147

                QRvaiARaLltkpkllLLDEPTagLDpasraqlllellrelrqgggTvlli
                +---t+      D+ +  +l ++lre ++
148  ----FMILIFTRK-----DDLGDNLHDYLRAP----- 173

                tHdldlldrla.DrilvledG<-*
                d+++l+++ +Dr + l++
174  --DIQDLMDIFgDRYCALNNK      192

```

FIGURE 8

GTCGACCCACGCGTCCGCCCCGATGCCCGGGCCCCGAGGGGCTGCTGGCGGCCTGGCCCCTGAG																			M	R	G	3
																			ATG	CGC	GGG	73
A	G	A	A	G	L	L	A	L	L	L	L	L	L	L	L	L	L	G	L	23		
CGC	GGG	GCG	GCG	GGG	CTG	CTG	GCG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	GGC	CTG	133		
G	G	R	V	E	G	G	P	A	G	E	R	G	A	G	G	G	G	A	L	43		
GGC	GGC	AGG	GTC	GAG	GGG	GGG	CCG	GCC	GGC	GAG	CGG	GGC	GCA	GGC	GGG	GGC	GGG	GCG	CTG	193		
A	R	E	R	F	K	V	V	F	A	P	V	I	C	K	R	T	C	L	K	63		
CCC	CGC	GAG	CGC	TTC	AAG	GTG	GTC	TTT	GCG	CCG	GTG	ATC	TGC	AAG	CGG	ACC	TGT	CTC	AAG	253		
G	Q	C	R	D	S	C	Q	Q	G	S	N	M	T	L	I	G	E	N	G	83		
GGC	CAG	TGT	CGG	GAC	AGT	TGT	CAG	CAG	GGC	TCC	AAC	ATG	ACG	CTC	ATC	GGA	GAG	AAC	GGC	313		
H	S	T	D	T	L	T	G	S	G	F	R	V	V	V	C	P	L	P	C	103		
CAC	AGC	ACA	GAC	ACG	CTC	ACG	GGC	TCC	GGC	TTC	CGC	GTG	GTG	GTG	TGC	CCT	CTC	CCC	TGC	373		
M	N	G	G	Q	C	S	S	R	N	Q	C	L	C	P	P	D	F	T	G	123		
ATG	AAT	GGC	GGC	CAG	TGC	TCC	TCG	CGA	AAC	CAG	TGC	CTG	TGT	CCC	CCG	GAC	TTC	ACT	GGG	433		
R	F	C	Q	V	P	A	G	G	A	G	G	G	T	G	G	S	G	P	G	143		
CGC	TTC	TGC	CAG	GTG	CCC	GCA	GGA	GGA	GCC	GGT	GGG	GGT	ACC	GGC	GGC	TCA	GGC	CCC	GGC	493		
L	S	R	T	G	A	L	S	T	G	A	L	P	P	L	A	P	E	G	D	163		
CTG	AGC	AGG	ACA	GGG	GCC	CTG	TCC	ACA	GGG	GCG	CTG	CCG	CCC	CTG	GCT	CCG	GAG	GGC	GAC	553		
S	V	A	S	K	H	A	I	Y	A	V	Q	V	I	A	D	P	P	G	P	183		
TCT	GTG	GCC	AGC	AAG	CAC	GCC	ATC	TAC	GCC	GTC	CAG	GTG	ATC	GCT	GAC	CCT	CCT	GGG	CCC	613		
G	E	G	P	P	A	Q	H	A	A	F	L	V	P	L	G	P	G	Q	I	203		
GGG	GAG	GGG	CCT	CCT	GCC	CAG	CAC	GCA	GCC	TTC	CTG	GTG	CCC	CTA	GGC	CCG	GGA	CAG	ATC	673		
S	A	E	V	Q	A	P	P	P	V	V	N	V	R	V	H	H	P	P	E	223		
TCA	GCA	GAA	GTG	CAG	GCC	CCG	CCC	CCC	GTG	GTG	AAT	GTG	CGC	GTC	CAT	CAC	CCG	CCC	GAG	733		
A	S	V	Q	V	H	R	I	E	S	S	N	A	E	S	A	A	P	S	Q	243		
CCC	TCA	GTC	CAG	GTG	CAC	CGC	ATT	GAG	AGC	TCG	AAC	GCC	GAG	AGC	GCA	GCC	CCC	TCC	CAG	793		
H	L	L	P	H	P	K	P	S	H	P	R	P	P	T	Q	K	P	L	G	263		
CAC	CTG	CTG	CCG	CAC	CCC	AAG	CCC	TCG	CAC	CCT	CGG	CCG	CCC	ACC	CAG	AAG	CCC	CTG	GGC	853		
R	C	F	Q	D	T	L	P	K	Q	P	C	G	S	N	P	L	P	G	L	283		
CGC	TGC	TTT	CAG	GAC	ACT	CTG	CCC	AAG	CAG	CCG	TGT	GGC	AGC	AAC	CCC	CTC	CCC	GGC	CTC	913		
T	K	Q	E	D	C	C	G	S	I	G	T	A	W	G	Q	S	K	C	H	303		
ACC	AAG	CAG	GAA	GAC	TGC	TGC	GGT	AGC	ATC	GGC	ACT	GCC	TGG	GGC	CAG	AGC	AAG	TGC	CAC	973		
K	C	P	Q	L	Q	Y	T	G	V	Q	K	P	G	P	V	R	G	E	V	323		
AAG	TGT	CCC	CAG	CTG	CAG	TAC	ACA	GGA	GTG	CAG	AAG	CCA	GGG	CCT	GTA	CGT	GGG	GAA	GTG	1033		
G	A	D	C	P	Q	G	Y	K	R	L	N	S	T	H	C	Q	D	I	N	343		
GGC	GCT	GAC	TGT	CCC	CAG	GGC	TAC	AAG	AGG	CTT	AAC	AGC	ACC	CAC	TGC	CAG	GAC	ATC	AAC	1093		
E	C	A	M	P	G	V	C	R	H	G	D	C	L	N	N	P	G	S	Y	363		
GAG	TGC	GCA	ATG	CCG	GGC	GTG	TGT	CGC	CAT	GGT	GAC	TGC	CTC	AAC	AAC	CCT	GGC	TCC	TAT	1153		

FIGURE 9A

R	C	V	C	P	P	G	H	S	L	G	P	S	R	T	Q	C	I	A	D	383
CGC	TGT	GTC	TGC	CCA	CCT	GGC	CAT	AGT	TTA	GGC	CCC	TCC	CGT	ACA	CAG	TGC	ATT	GCA	GAC	1213
K	P	E	E	K	S	L	C	F	R	L	V	S	P	E	H	Q	C	Q	H	403
AAA	CCG	GAG	GAG	AAG	AGC	CTG	TGT	TTC	CGC	CTG	GTG	AGC	CCT	GAG	CAC	CAG	TGC	CAG	CAC	1273
P	L	T	T	R	L	T	R	Q	L	C	C	C	S	V	G	K	A	W	G	423
CCA	CTG	ACC	ACC	CGC	CTG	ACC	CGC	CAG	CTC	TGC	TGC	TGC	AGT	GTC	GGC	AAG	GCC	TGG	GGC	1333
A	R	C	Q	R	C	P	T	D	G	T	A	A	F	K	E	I	C	P	A	443
GCG	CGG	TGT	CAG	CGC	TGC	CCA	ACA	GAT	GGC	ACC	GCT	GCG	TTC	AAG	GAG	ATC	TGC	CCA	GCT	1393
G	K	G	Y	H	I	L	T	S	H	Q	T	L	T	I	Q	G	E	S	D	463
GGG	AAG	GGA	TAC	CAC	ATT	CTC	ACC	TCC	CAC	CAG	ACG	CTC	ACC	ATT	CAG	GGC	GAG	AGT	GAC	1453
F	S	L	F	L	H	P	D	G	P	P	K	P	Q	Q	L	P	E	S	P	483
TTT	TCC	CTT	TTC	CTG	CAC	CCT	GAC	GGG	CCA	CCC	AAG	CCC	CAG	CAG	CTT	CCG	GAG	AGC	CCT	1513
S	Q	A	P	P	P	E	D	T	E	E	E	R	G	V	T	T	D	S	P	503
AGC	CAG	GCT	CCA	CCA	CCT	GAG	GAC	ACA	GAG	GAA	GAG	AGA	GGG	GTG	ACC	ACG	GAC	TCA	CCG	1573
V	S	E	E	R	S	V	Q	Q	S	H	P	T	A	T	T	T	P	A	R	523
GTG	AGT	GAG	GAG	AGG	TCA	GTG	CAG	CAG	AGC	CAC	CCA	ACT	GCC	ACC	ACG	ACT	CCT	GCC	CGG	1633
P	Y	P	E	L	I	S	R	P	S	P	P	T	M	R	W	F	L	P	D	543
CCC	TAC	CCC	GAG	CTG	ATC	TCC	CGT	CCC	TCG	CCC	CCG	ACC	ATG	CGC	TGG	TTC	CTG	CCG	GAC	1693
L	P	P	S	R	S	A	V	E	I	A	P	T	Q	V	T	E	T	D	E	563
TTG	CCT	CCT	TCC	CGC	AGC	GCC	GTA	GAG	ATC	GCT	CCC	ACT	CAG	GTC	ACA	GAG	ACT	GAT	GAG	1753
C	R	L	N	Q	N	I	C	G	H	G	E	C	V	P	G	P	P	D	Y	583
TGC	CGA	CTG	AAC	CAG	AAC	ATC	TGT	GGC	CAC	GGA	GAG	TGC	GTG	CCG	GGC	CCC	CCT	GAC	TAC	1813
S	C	H	C	N	P	G	Y	R	S	H	P	Q	H	R	Y	C	V	D	V	603
TCC	TGC	CAC	TGC	AAC	CCC	GGC	TAC	CGG	TCA	CAT	CCC	CAG	CAC	CGC	TAC	TGC	GTG	GAT	GTG	1873
N	E	C	E	A	E	P	C	G	P	G	R	G	I	C	M	N	T	G	G	623
AAC	GAG	TGC	GAG	GCA	GAG	CCC	TGT	GGC	CCG	GGG	AGG	GGC	ATC	TGC	ATG	AAC	ACC	GGC	GGC	1933
S	Y	N	C	H	C	N	R	G	Y	R	L	H	V	G	A	G	G	R	S	643
TCC	TAC	AAT	TGC	CAC	TGC	AAC	CGC	GGC	TAC	CGC	CTG	CAC	GTG	GGC	GCC	GGG	GGG	CGC	TCG	1993
C	V	D	L	N	E	C	A	K	P	H	L	C	G	D	G	G	F	C	I	663
TGC	GTG	GAC	CTG	AAC	GAA	TGC	GCC	AAG	CCC	CAC	CTG	TGC	GGC	GAC	GGC	GGC	TTC	TGC	ATC	2053
N	F	P	G	H	Y	K	C	N	C	Y	P	G	Y	R	L	K	A	S	R	683
AAC	TTT	CCC	GGT	CAC	TAC	AAG	TGC	AAC	TGC	TAC	CCC	GGC	TAC	CGG	CTC	AAA	GCC	TCC	CGG	2113
P	P	V	C	E	D	I	D	E	C	R	D	P	S	S	C	P	D	G	K	703
CCT	CCT	GTG	TGC	GAA	GAC	ATC	GAC	GAG	TGC	CGG	GAC	CCA	AGC	TCT	TGC	CCG	GAT	GGC	AAA	2173
C	E	N	K	P	G	S	F	K	C	I	A	C	Q	P	G	Y	R	S	Q	723
TGC	GAG	AAC	AAG	CCC	GGG	AGC	TTC	AAG	TGC	ATC	GCC	TGT	CAG	CCT	GGC	TAC	CGC	AGC	CAG	2233
G	G	G	A	C	R	D	V	N	E	C	A	E	G	S	P	C	S	P	G	743
GGG	GGC	GGG	GCC	TGT	CGC	GAC	GTG	AAC	GAG	TGC	GCC	GAG	GGC	AGC	CCC	TGC	TCG	CCT	GGC	2293
W	C	E	N	L	P	G	S	F	R	C	T	C	A	Q	G	Y	A	P	A	763
TGG	TGC	GAG	AAC	CTC	CCG	GGC	TCC	TTC	CGC	TGC	ACC	TGT	GCC	CAG	GGC	TAC	GCG	CCC	GCG	2353

FIGURE 9B

P	D	G	R	S	C	L	D	V	D	E	C	E	A	G	D	V	C	D	N	783
CCC	GAC	GGC	CGC	AGT	TGC	TTG	GAT	GTG	GAC	GAG	TGT	GAG	GCT	GGG	GAC	GTG	TGT	GAC	AAT	2413
G	I	C	S	N	T	P	G	S	F	Q	C	Q	C	L	S	G	Y	H	L	803
GGC	ATC	TGC	AGC	AAC	ACG	CCA	GGA	TCT	TTC	CAG	TGT	CAG	TGC	CTC	TCT	GGC	TAC	CAT	CTG	2473
S	R	D	R	S	H	C	E	D	I	D	E	C	D	F	P	A	A	C	I	823
TCC	AGG	GAC	CGG	AGC	CAC	TGC	GAG	GAC	ATT	GAT	GAG	TGT	GAC	TTC	CCT	GCA	GCC	TGC	ATT	2533
G	G	D	C	I	N	T	N	G	S	Y	R	C	L	C	P	Q	G	H	R	843
GGG	GGT	GAC	TGC	ATC	AAT	ACC	AAT	GGC	TCC	TAC	AGA	TGT	CTT	TGC	CCC	CAG	GGG	CAT	CGG	2593
L	V	G	G	R	K	C	Q	D	I	D	E	C	S	Q	D	P	S	L	C	863
CTG	GTG	GGT	GGC	AGG	AAA	TGC	CAA	GAC	ATA	GAT	GAG	TGC	AGC	CAG	GAC	CCG	AGC	CTG	TGC	2653
L	P	H	G	A	C	K	N	L	Q	G	S	Y	V	C	V	C	D	E	G	883
CTT	CCC	CAT	GGG	GCC	TGC	AAG	AAC	CTT	CAG	GGC	TCC	TAT	GTG	TGT	GTC	TGC	GAT	GAG	GGC	2713
F	T	P	T	Q	D	Q	H	G	C	E	E	V	E	Q	P	H	H	K	K	903
TTC	ACT	CCC	ACC	CAG	GAC	CAG	CAC	GGT	TGT	GAG	GAG	GTG	GAG	CAG	CCC	CAC	CAC	AAG	AAG	2773
E	C	Y	L	N	F	D	D	T	V	F	C	D	S	V	L	A	T	N	V	923
GAG	TGC	TAC	CTG	AAC	TTC	GAT	GAC	ACA	GTG	TTC	TGC	GAC	AGC	GTA	TTG	GCC	ACC	AAC	GTG	2833
T	Q	Q	E	C	C	C	S	L	G	A	G	W	G	D	H	C	E	I	Y	943
ACC	CAG	CAG	GAG	TGC	TGC	TGC	TCT	CTG	GGG	GCC	GGC	TGG	GGC	GAC	CAC	TGC	GAA	ATC	TAC	2893
P	C	P	V	Y	S	S	A	E	F	H	S	L	C	P	D	G	K	G	Y	963
CCC	TGC	CCA	GTC	TAC	AGC	TCA	GCC	GAG	TTC	CAC	AGC	CTC	TGC	CCA	GAC	GGA	AAG	GGC	TAC	2953
T	Q	D	N	N	I	V	N	Y	G	I	P	A	H	R	D	I	D	E	C	983
ACC	CAG	GAC	AAC	AAC	ATC	GTC	AAC	TAC	GGC	ATC	CCA	GCC	CAC	CGT	GAC	ATC	GAC	GAG	TGC	3013
M	L	F	G	S	E	I	C	K	E	G	K	C	V	N	T	Q	P	G	Y	1003
ATG	TTG	TTC	GGG	TCG	GAG	ATT	TGC	AAG	GAG	GGC	AAG	TGC	GTG	AAC	ACG	CAG	CCT	GGC	TAC	3073
E	C	Y	C	K	Q	G	F	Y	Y	D	G	N	L	L	E	C	V	D	V	1023
GAG	TGC	TAC	TGC	AAG	CAG	GGC	TTC	TAC	TAC	GAC	GGG	AAC	CTG	CTG	GAA	TGC	GTG	GAC	GTG	3133
D	E	C	L	D	E	S	N	C	R	N	G	V	C	E	N	T	R	G	G	1043
GAC	GAG	TGC	CTG	GAC	GAG	TCC	AAC	TGC	CGG	AAC	GGA	GTG	TGT	GAG	AAC	ACG	CGC	GGC	GGC	3193
Y	R	C	A	C	T	P	P	A	E	Y	S	P	A	Q	R	Q	C	L	S	1063
TAC	CGC	TGT	GCC	TGC	ACG	CCC	CCT	GCC	GAG	TAC	AGT	CCC	GCG	CAG	CGC	CAG	TGC	CTG	AGC	3253
P	E	EG	M	D	V	D	E	C	Q	D	P	A	A	C	R	P	G	R	C	1083
CCG	GAA	GAG	ATG	GAC	GTG	GAC	GAG	TGC	CAG	GAC	CCG	GCA	GCC	TGC	CGC	CCT	GGC	CGC	TGC	3313
V	N	L	P	G	S	Y	R	C	E	C	R	P	P	W	V	P	G	P	S	1103
GTC	AAC	CTG	CCG	GGC	TCC	TAC	CGC	TGC	GAG	TGT	CGC	CCG	CCC	TGG	GTG	CCC	GGG	CCC	TCC	3373
G	R	D	C	Q	L	P	E	S	P	A	E	R	A	P	E	R	R	D	V	1123
GGC	CGC	GAT	TGC	CAG	CTC	CCC	GAG	AGC	CCG	GCC	GAG	CGT	GCC	CCG	GAG	CGG	CGC	GAC	GTG	3433
C	W	S	Q	R	G	E	D	G	M	C	A	G	P	L	A	G	P	A	L	1143
TGC	TGG	AGC	CAG	CGC	GGA	GAG	GAC	GGC	ATG	TGC	GCT	GGC	CCC	CTG	GCC	GGG	CCT	GCC	CTC	3493
T	F	D	D	C	C	C	R	Q	G	R	G	W	G	A	Q	C	R	P	C	1163
ACC	TTC	GAC	GAC	TGC	TGC	TGC	CGC	CAG	GGC	CGC	GGC	TGG	GGC	GCC	CAA	TGC	CGA	CCG	TGC	3553

FIGURE 9C

P	P	R	G	A	G	S	H	C	P	T	S	Q	S	E	S	N	S	F	W	1183
CCG	CCG	CGC	GGC	GCG	GGG	TCC	CAT	TGC	CCG	ACA	TCG	CAG	AGC	GAG	AGC	AAT	TCC	TTC	TGG	3613
D	T	S	P	L	L	L	G	K	P	P	R	D	E	D	S	S	E	E	D	1203
GAC	ACA	AGC	CCC	CTG	CTG	TTG	GGG	AAG	CCC	CCA	AGA	GAT	GAG	GAC	AGT	TCA	GAG	GAG	GAT	3673
S	D	E	C	R	C	V	S	G	R	C	V	P	R	P	G	G	A	A	C	1223
TCA	GAC	GAG	TGT	CGC	TGC	GTG	AGT	GGC	CGC	TGC	GTG	CCG	CGG	CCG	GGC	GGC	GCC	GCG	TGC	3733
E	C	P	G	G	F	Q	L	D	A	S	R	A	R	C	V	D	I	D	E	1243
GAG	TGT	CCC	GGC	GGC	TTC	CAG	CTC	GAC	GCC	TCC	CGC	GCC	CGC	TGC	GTG	GAT	ATC	GAC	GAG	3793
C	R	E	L	N	Q	R	G	L	L	C	K	S	E	R	C	V	N	T	S	1263
TGC	CGA	GAG	CTG	AAC	CAG	CGC	GGG	CTG	CTG	TGC	AAG	AGC	GAG	CGC	TGC	GTG	AAC	ACC	AGC	3853
G	S	F	R	C	V	C	K	A	G	F	A	R	S	R	P	H	G	A	C	1283
GGC	TCC	TTC	CGC	TGC	GTC	TGC	AAA	GCC	GGC	TTC	GCG	CGC	AGC	CGC	CCG	CAC	GGG	GCC	TGC	3913
V	P	Q	R	R	R	*														1290
GTT	CCC	CAG	CGC	CGC	CGC	TGA														3934
CGCCGCCGACGCCGCCCTCGGCCAGACCTCGGTGATCACTGAGGGATTTCCGCGAGCTCGGCCTCACTTCTGCCCCGA	4013																			
CTTGTGGCTCGGACCCAGGGACCTTCAGGGCCCGCAGACCCTCCCGGCGCCTTGAGACCCGAGGCGCCCCCTACCGGCCC	4092																			
CCCTCCCCGGTTAGCGGGCGGTTGTAAGGTCTCCGGCGGGCGCTGCCTGCCTTCCTCCAGAGGGTGTTCCTAGAAAC	4171																			
TGATAAATCAGATCGTGCCTCTTTAAAAAAAAAAAAAAAAAAAAAGGGCGGCCGC	4225																			

FIGURE 9D

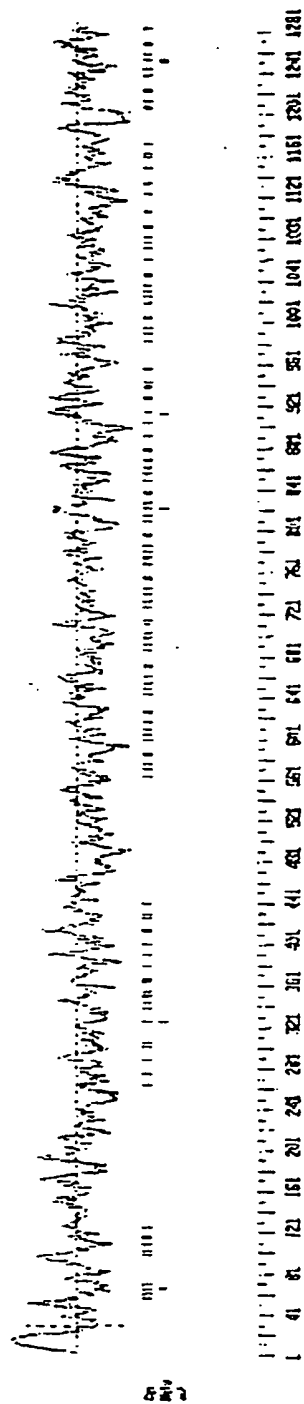


FIGURE 10

hT275 *->CnpntgpcInggGtCvntpggsvfggytCeCpeGyalsytGkrC<-
 C++ pC+ngG+C +C Cp+ +tG++C
 CPL---pCMNGGQCSSRN-----QCLCPPD-----FTGRFC

hT275 *->CnpntgpcInggGtCvntpggsvfggytCeCpeGyalsytGkrC<-
 C++++ +C + G C+n pg +y+C+Cp+G++l+ + +C
 CAMPG-VCRH-GDCLNMPG-----SYRCVCPGHSGLPSRTQC

hT275 *->CnpntgpcInggGtCvntpggsvfggytCeCpeGyalsytGkrC<-
 C -n + C + G+Cv++p +y+C+C +Gy+ + + +C
 CRLNQNICGH-GECVPGPP-----DYSCHCNPGYRSHPOHRYC

hT275 *->CnpntgpcInggGtCvntpggsvfggytCeCpeGyals..ytGkrC<-
 C + pC g+G+C+nt+g +y C+C Gy+l+ + G+ C
 CEAE--pCGPGrGICMNTGG-----SYNCHCNRGYRLHvgAGGRSC

hT275 *->CnpntgpcInggGtCvntpggsvfggytCeCpeGyals..ytGkrC<-
 C++++ C gG+C+n pg +y C C +Gy+l+ + C
 CAKPH-LCGDGGFCINFPG-----HYKCNCYPGYRLKaSRPPVC

hT275 *->CnpntgpcInggGtCvntpggsvfggytCe.CpeGyalsytGkrC<-
 C ++ C G+C n pg ++ C+ C++Gy+ G C
 CRDPS-SCPD-GKCENKPG-----SFKCIAcQPGYRS-QGGGAC

hT275 *->CnpntgpcInggGtCvntpggsvfggytCeCpeGyalsytGkrC<-
 C+ + pC+ G C n+pg +++C+C++Gya+ +G+ C
 CAEGS-PCSP-GWCENLPG-----SFRCTCAQGYAPAPDGRSC

hT275 *->CnpntgpcInggGtCvntpggsvfggytCeCpeGyalsytGkrC<-
 C ++ +C+n G+C ntpg +++C+C Gy ls + +C
 CEAGD-VCDN-GICSNTPG-----SFQCQCLSGYHLSRDRSHC

hT275 *->CnpntgpcInggGtCvntpggsvfggytCeCpeGyalsytGkrC<-
 C+ ++ C + G C+nt g +y+C Cp+G++l G++C
 CDFPA-ACIG-GDCINTNG-----SYRCLCPQGHRL-VGGRKC

hT275 *->CnpntgpcInggGtCvntpggsvfggytCeCpeGyalsytGkrC<-
 C++++ Cl +G C n++g +y C+C+eG+ + + +C
 CSQDPSLCLPHGACKNLQG-----SYVCVCDEGFTPTQDQHGC

hT275 *->Cnpntg.pCInggGtCvntpggsvfggytCeCpeGyalsytGkrC<-
 C + + + C+ G+Cvnt+ gy+C C++G+++ + C
 CMLFGSeICKE-GKCVNTQP-----GYECYCKQGFYDGNLLEC

FIGURE 11A

```

      *-->CnpntgpcInGtCvntpggsvfggytCeCpeGyals.ytGkrC<--*
      C **  +C n G C nt g      gy+C C+++ a+++ + ++C
ht275      CLDES-NCRN-GVCENTRG-----GYRCACTPP-AEYsPAQRQ

      *-->CnpntgpcInGtCvntpggsvfggytCeCpeGyalsytGkrC<--*
      C++++ C  G+Cvn+pg      +y+CeC **  ++ +G++C
ht275      CQDPA-ACRP-GRCVNLPG-----SYRCECRPPWVPGPSGRDC

      *-->CnpntgpcInGtCvntpggsvfggytCeCpeGyalsytGkrC<--*
      *** + C  G+Cv+ pg      g CeCp G++l + rC
ht275      -DSDECRCVS-GRCVPRPG-----GAACECPGGFQLDASRARC

      *-->Cnpntg...pCInGtCvntpggsvfggytCeCpeGyalsytGkrC<
      C ** ++ C+  +Cvnt g      +++C+C+ G+a s      C
ht275      CRELNQrglLCKSE-RCVNTSG-----SFRVCVKAGFARSRPHGAC

```

FIGURE 11B

hT275 *->grC.snplpgravTKse.CCCsvGrgeAWGtp.CElCPvpgtaefke
 + C+snplpg +TK+e+CC s G+ AWG +C +CP + + ++
 QPCgSNPLPG--LTKQEdCCGSIGT--AWGQSkCHKCPQLQYTGvQK 329

L<-*

hT275 .. P~ 330

hT275 *->grCsnplpgravTKseCCCsvGrgeAWGtpCElCPvpgtaefkeL<-
 +C++pl r +T++ CCCsvG+ AWG C+ CP++gta+fke+
 HQCQHPLTTR-LTRQLCCCSVGK--AWGARCQRCPTDGTAAFKEI

hT275 *->grCsnplpgravTKseCCCsvGrgeAWGtpCE..lCPvpgtaefkeL
 C+ l+ + vT +eCCCS+G+ +WG++CE +CPv +aef+ L
 VFCDSVLATN-VTQQECCCSLGA--GWGDHCElyPCPVYSSAEFHSL

<-*

hT275 - -

hT275 *->grCsnplpgravTKseCCCsvGrgeAWGtpCElCPvpg...taefke
 g+C +pl+g a+T + CCC Gr +WG +C +CP++g +++++ +
 GMCAGPLAGPALTFDDCCCRQGR--GWGAQCRPCPPRGagsHCPTSQ

L<-*

hT275 E

FIGURE 12

```

*->mDPqnCsCatggsCtCgtsCkCknC.kCtsCkKsccsCcPagCskCa
      DP   sC+ g   +C+++   +C C++ +s + + + +Ca
hT275      RDP--SSCPDG---KCENKPGSFKCiACQPGYRSQGGGACRDVNECA

      qqCvCkgg..gaasetSkCsCCa<--
      +g C g ++ ++ C C++
hT275      EGSPCSPGwcENLPGSFRCTCAQ

```

FIGURE 13

ALIGN calculates a global alignment of two sequences

version 2.0uPlease cite: Myers and Miller, CABIOS (1989)

> hT275 n.a.

4225 aa vs.

> L40459 n.a.

4317 aa

scoring matrix: pam120.mat, gap penalties: -12/-4

77.1% identity; Global alignment score: 16879

```

                                10          20
inputs  -----GTC-----GACCCA-----CGCGTCC-----GCCCCG--GATGCC-----
          :::          ::::  :  ::  :  ::::  :  :::
          CCTCCTGCTGTCCCCTCCCTACCCTTGGCTTCTCGCCCCGCTCTGCCCTCTGCTACCAACACTCGATCCC
          10          20          30          40          50          60          70

          30          40          50          60          70
inputs  -----CGGGC-C-----CCGAGGGGC-TGCTGGC--GG-----CCTGGCCCC-TGAGATGCGCG
          ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  ::::  :
          CTGCTCGGGCTCGACCTCCAATCTCCGAGGGTCTGTCGGCCCCCGGATGCCCGGGCCCCGAGCGGTGCCCA
          80          90          100          110          120          130          140

          80          90          100          110          120          130
inputs  GGGCGGGGGC--GGCGGGGCTGCTGGCGCTGCTGCT---GCTGCTGCTGCTGCTGCTGCTGG-GC--CT
          :::  ::  :  ::::  ::  ::::  :  :::  :  ::::  :  ::::  :  ::::  :  :::
          CGGCCTGGCCCCCTGCGATGCGCCAGGC-CGGCGGATTGGGGCTGCTGGCACTACTCCTGCTGGCGCTGCT
          150          160          170          180          190          200

          140          150          160          170          180          190          200
inputs  GGGCGGCAG-GGTTCGAGGGG-GGGCCGGCCGGCGAGCGGGGCGCAGGCGGGGGCGGGGCGCTGGCCCCGC
          ::::  ::  :  ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  :::
          GGGCCCCGGCGGCCGAGGGGTGGGCGGCCGGGCAGCGGGGCACAGGCGGGGGCGGGGCGCTGGGCC-C-
          210          220          230          240          250          260          270

          210          220          230          240          250          260          270
inputs  AGCGCTTCAAGGTGGTCTTTGCGCCGGTGATCTGCAAGCGGACCTGTCTCAAGGGCCAGTGTCTGGGACAG
          ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  ::::  :
          AACGCTTCAAGGTGGTCTTTGCGCCTGTGATCTGCAAGCGGACCTGTCTGAAGGGCCAGTGTCTGGGACAG
          280          290          300          310          320          330          340

          280          290          300          310          320          330          340
inputs  TTGTCAGCAGGGCTCCAACATGACGCTCATCGGAGAGAACGGCCACAGCACAGACACGCTCACGGGCTCC
          ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  :::
          CTGTCAGCAGGGCTCCAACATGACGCTCATCGGAGAGAACGGCCACAGCACCGACACGCTCACCGGTTCT
          350          360          370          380          390          400          410

          350          360          370          380          390          400          410
inputs  GGCTTCCGCGTGGTGGTGTGCCCTCTCCCTGCGTGAATGGCGGCCAGTGCTCCTCGCGAAACCAGTGCC
          :  ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  :::
          GCCTTCCGCGTGGTGGTGTGCCCTCTACCCTGCGTGAACGGTGGCCAGTGCTCTTCCCGAAACCAGTGCC
          420          430          440          450          460          470          480

          420          430          440          450          460          470          480
inputs  TGTGTCCCCCGGACTTCACTGGGCGCTTCTGCCAGGTGCCCGCAGGAGGAGCCGGTGGGGGTACCGGCGG
          ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  ::::  :  :::
          TGTGTCCCCCGGATTTCACGGGGCGCTTCTGCCAGGTGCCCTGCTGCAGGAACCGGAGCTGGCACCGGGAG
          490          500          510          520          530          540          550

```

FIGURE 14A

[illegible]

21/60

FIGURE 14 D

23/60

```

      3000      3010      3020      3030      3040      3050      3060
inputs  CACCGTGACATCGACGAGTGCATGTTGTTTCGGGTCGGAGATTGTGCAAGGAGGGCAAGTGCCTGAACACGC
      ..... : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      CACCGTGACATCGACGAATGCATATTGTTTGGGGCAGAGATCTGCAAGGAGGGCAAGTGTGTGAACACGC
      3070      3080      3090      3100      3110      3120      3130

      3070      3080      3090      3100      3110      3120      3130
inputs  AGCCTGGCTACGAGTGCTACTGCAAGCAGGGCTTCTACTACGACGGGAACCTGCTGGAATGCGTGACGCT
      .... : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      AGCCCGGCTACGAGTGCTACTGCAAGCAGGGCTTCTACTACGATGGCAACCTGCTGGAGTGCCTGACGCT
      3140      3150      3160      3170      3180      3190      3200

      3140      3150      3160      3170      3180      3190      3200
inputs  GGACGAGTGCCTGGACGAGTCCAAGTCCCGGAACGGAGTGTGTGAGAACACGCGCGGGCTACCGCTGT
      ... : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      GGATGAGTGCCTGGATGAGTCTAACTGCAGGAACGGAGTGTGTGAGAACACACGCTGGCGGGCTACCGCTGT
      3210      3220      3230      3240      3250      3260      3270

      3210      3220      3230      3240      3250      3260      3270
inputs  GCCTGCACGCCCCCTGCCGAGTACAGTCCCGCGCAGCGCCAGTGCCTGAGCCCGGAAGAGATGGACGCTGG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      GCCTGCACTCCGCCGGCAGAGTACAGTCCCGCACAGGCCAGTGTCTGATCCCGGA-GAGATGGA-----
      3280      3290      3300      3310      3320      3330      3340

      3280      3290      3300      3310      3320      3330      3340
inputs  ACGAGTGCCAGGACCCGGCAGCCTGCCGCCCTGGCCGCTGCGTCAACCTGCCGGGCTCCTACCGCTGCGA
      ..... : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      -----GCAC-----

      3350      3360      3370      3380      3390      3400      3410
inputs  GTGTCGCCCCCTGGGTGCCCGGGGCCCTCCGGCCGCGATTGCCAGCTCCCCGAGAGCCCGGCCGAGCGT
      -----

      3420      3430      3440      3450      3460      3470      3480
inputs  GCCCCGAGCGGCGCGACGTGTGCTGGAGCCAGCGCGGAGAGGACGGCATGTGCGCTGGCCCCCTGGCCG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      GCCCCAGAGAGACGTGAAGTGTGCTGGGGCCAGCGAGGAGAGGACGGCATGTGTATGGGGCCCTGGGGCG
      3350      3360      3370      3380      3390      3400      3410

      3490      3500      3510      3520      3530      3540      3550
inputs  GG-CCTGCCCTCACCTTCGACGACTGCTGCTGCCGCCAGGGCCGCGGCTGGGGCGCCCAATGCCGACCGT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      GGACCTGCCCTCACTTTTGATGACTGCTGCTGCCGCCAG--CCGCGGCTGGGT-ACCCAGTGACGACCGT
      3420      3430      3440      3450      3460      3470      3480

      3560      3570      3580      3590      3600      3610      3620
inputs  GCCCCGCCGCGCGGCGGGGTCCCATGCCCCGACATCGCAGAGCGAGAGCAATTCTTCTGGGACACAAG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      GCCCCCACGTGGCACC GGTTCCAGTGCCGACTTCACAGAGTGAGAGCAATTCTTCTGGGACACAAG
      3490      3500      3510      3520      3530      3540      3550

```

FIGURE 14F

```

      3630      3640      3650      3660      3670      3680      3690
inputs  CCCCCTGCTGTTGGGGAAGCCCCCAAGAGATGAGGACAGTTCAGAGGAGGATTTCAGACGAGTGTGCTGCTGC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      CCCCCTGCTACTGGGGAAGTCTCCGCGAGACGAAGACAGCTCAGAGGAGGATTTCAGATGAGTGCCGTTGT
      3560      3570      3580      3590      3600      3610      3620

      3700      3710      3720      3730      3740      3750      3760
inputs  GTGAGTGGCCGCTGCGTGCCGCGGCCGGGCGGCCGCGTGCAGTGTCCCGGCGGCTTCAGCTCGACG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      GTGAGCGGACGCTGTGTGCCACGGCCAGGCGGGGCGGTATGCGAGTGTCTCTGGAGGCTTTTCAGCTGGACG
      3630      3640      3650      3660      3670      3680      3690

      3770      3780      3790      3800      3810      3820      3830
inputs  CCTCCCGCGCCCGCTGCGTGGATATCGACGAGTGCCGAGAGCTGAACCAGCGCGGGCTGCTGTGCAAGAG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      CCTCCCGTGCCCGCTGCGTGGACATTGATGAGTGCCGAGAACTGAACCAGCGGGGACTGCTGTGTAAGAG
      3700      3710      3720      3730      3740      3750      3760

      3840      3850      3860      3870      3880      3890      3900
inputs  CGAGCGCTGCGTGAACACCAGCGGCTCCTTCCGCTGCGTCTGCAAAGCCGGCTTCGCGCGCAGCCGCCCCG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      CGAGCGGTGCGTGAACACCAGTGGATCCTTCCGCTGTGTCTGCAAAGCTGGCTTCACGCGCAGCCGCCCCCT
      3770      3780      3790      3800      3810      3820      3830

      3910      3920      3930      3940      3950      3960
inputs  CACGGGGCCTGCGTTCCCCAGCGCCCGCGCTGACG---CCGCCGACGCC-GC-CCTCGGC--CCAGACCT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      CACGGG-CCTGCGTGCCCTCAGCGCCCGCGCTGATGATGCAGCCATAGCCCACACCTCAGTGATCGATCAT
      3840      3850      3860      3870      3880      3890      3900

      3970      3980      3990      4000      4010      4020
inputs  CGGTGAT-----CACTGA--GGGATTTCCG-CGAGCTCGGCCTCA-CTTCTGCCCCGACTTGTGGCTCGG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      CGAGGGTATTTTCACTGAAAGTGGAGACAGACAAGTACATCCTTTGCTCCTGACCAAACGAGAG-CATGG
      3910      3920      3930      3940      3950      3960

      4030      4040      4050      4060      4070      4080      4090
inputs  ACCCAGGGA-CCTTCAGGGCCCGCAGA-C-CCTCCCGGCGCCTTGAGACCCGAGGCGCCCCCTACCGGC--
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      -ACCCAAGGATCCTTCAGGGCCCAAAATCTCCTTCCACACCCC-AAACCCAAGGTGCTCCTGTCTGCAG
      3970      3980      3990      4000      4010      4020      4030

                                                    4100
inputs  -----CCC-----CCT-----CCCC---GGTTAGC-
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      AGTGCTGTCTGCTTTCTCCCAAGGGTGATTCTAGAACTTCGACATCAGATCTGCCCCCTTAATTTACT
      4040      4050      4060      4070      4080      4090      4100

      4110      4120      4130      4140
inputs  -----GGGCGGTGTGAAGGTC---TCCG--GCGGGC-GCTGC--CTGCCT-----TC--
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      CTTGGCTTTCAAGGCAAATTGATATTACATCCAAAGCGGGCAGCATCAACTGCTTGGCGGGTTGGACTG
      4110      4120      4130      4140      4150      4160      4170

```

FIGURE 14G

```

      4150      4160      4170      4180      4190
inputs --CT---CCCAGAGGGTGT-----TTCCTAGAACT--GATAAA---TCAGATCGTGCCT---CTT
      ::      :::::::::: ::::      ::      :::::::::: :::::      :::::      :::::      ::
      AGCTGGGACCCAGGATGTGAAATAGAAATTTATTGTGGCTCTGATTATGTACACTAGATGTGCCTGACCTG
      4180      4190      4200      4210      4220      4230      4240

      4200      4210      4220
inputs -TAA-----AAAAAAAAAAAAAAAAAGGGCGGCCGC-----
      ::      : :::::::::: :::::      .:: ::::
      CTGACCAGGCTCACATGGTTTGTACAATAAATACATCCGCCGGGAAAAAAAAAAAAAAAAAAAAAAAAA
      4250      4260      4270      4280      4290      4300      4310
```

FIGURE 14H

ALIGN calculates a global alignment of two sequences

version 2.0uPlease cite: Myers and Miller, CABIOS (1989)

> Patent Protein R79475 - (untitled) 1251 aa vs.

> hT275 a.a. 1289 aa

scoring matrix: pam120.mat, gap penalties: -12/-4

82.8% identity; Global alignment score: 5580

```

      10      20      30      40      50      60
inputs MRQAA----LGLLALLLLALLGPGGRGVGRPGS--GAQAGAGRWAQRFKVVVFAPVICKRTCLKGQCRDSC
      :: ::  :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
      MRGAGAAGLLALLLLLLLLLLGLGGRVEGGPAGERGAGGGGALARERFKVVVFAPVICKRTCLKGQCRDSC
      10      20      30      40      50      60      70

      70      80      90      100      110      120      130
inputs QQGSNMTLIGENGHSTDTLTGSAFRVVVCPLPCMNGGQCSSRNQCLCPPDFTGRFCQVPAAGTGAGTGSS
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
      QQGSNMTLIGENGHSTDTLTGSGFRVVVCPLPCMNGGQCSSRNQCLCPPDFTGRFCQVPAAGGAGGGTGGS
      80      90      100      110      120      130      140

      140      150      160      170      180      190      200
inputs GPG-WPDRAMSTGPLPPLAPEGESVASKHAIYAVQVIADPPGPGEGPPAQHAAFLVPLGPGQISAEVQAP
      ::  :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
      GPGLSRTGALSTGALPPLAPEGDSVASKHAIYAVQVIADPPGPGEGPPAQHAAFLVPLGPGQISAEVQAP
      150      160      170      180      190      200      210

      210      220      230      240      250      260      270
inputs PPVVNVRVHHPPEASVQVHRIEGPNAEGPASSQHLLPHPKPQHPRPPTQKPLGRCFQDTLPKQPCGSNPL
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
      PPVVNVRVHHPPEASVQVHRIESSNAESAAPSQHLLPHPKPSHPRPPTQKPLGRCFQDTLPKQPCGSNPL
      220      230      240      250      260      270      280

      280      290      300      310      320      330      340
inputs PGLTKQEDCCGSIGTAWGQSKCHKCPQLQYTGVOQKVPVVRGEVGADCPQGYKRLNSTHCQDINECAMPGN
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
      PGLTKQEDCCGSIGTAWGQSKCHKCPQLQYTGVOQKPGPVRGEVGADCPQGYKRLNSTHCQDINECAMPGV
      290      300      310      320      330      340      350

      350      360      370      380      390      400      410
inputs VCHGDCLNNPGSYRCVCPPGHSLGPLAAQCIADKPPEEKSLCFRLVSTEHQCQHPLTTRLTRQLCCCSVGK
      :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
      CRHGDCLNNPGSYRCVCPPGHSLGPSRTQCIADKPPEEKSLCFRLVSPPEHQCQHPLTTRLTRQLCCCSVGK
      360      370      380      390      400      410      420

      420      430      440      450      460      470      480
inputs AWGARCQRCPADGTAAFKEICPGWE--RVYPHLPDPAHHPGGKRLPLPAPDGPPKQQLPESPSRAP
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
      AWGARCQRCPTDGTAAFKEICPAGKGYHILTSHTLTIQGESDFSLF-LH-PDGPPKQQLPESPSQAPP
      430      440      450      460      470      480

      490      500      510      520      530      540      550
inputs LEDTEEERGVTMDPPVSEERSVQQSHPTTTTSPPRYPPELISRPSPTTFHRFLPDLPPSRSAVEIAPTQV
      :::: :::: :::: :::: :::: :::: :::: :::: ::::
      PEDTEEERGVTDTSPVSEERSVQQSHPTATTTTPARYPPELISRPSPTTMRWFLPDLPPSRSAVEIAPTQV
      490      500      510      520      530      540      550

```

FIGURE 15A

```

      560      570      580      590      600      610      620
inputs TETDECRNLQNICGHGQCVPGPSDYSCHCNAGYRSHPQHRYCVDVNECEAEPCGPGKGICMNTGGSYNCH
      .....
      TETDECRNLQNICGHGECVPGPPDYSCHCNPGYRSHPQHRYCVDVNECEAEPCGPGRGICMNTGGSYNCH
      560      570      580      590      600      610      620

      630      640      650      660      670      680      690
inputs CNRGYRLHVGAGGRSCVDLNECAKPHLCGDDGGFCINFPGHYKNCNYPGYRLKASRPPICEDIDECRDPST
      .....
      CNRGYRLHVGAGGRSCVDLNECAKPHLCGDDGGFCINFPGHYKNCNYPGYRLKASRPPVCEDIDECRDPSS
      630      640      650      660      670      680      690

      700      710      720      730      740      750      760
inputs CPDGKCENTKPGSFKCIACQPGYRSQGGGACRDVNECSEGTSPGWCCEKLPGSYRCTCAQG-IRTRTGRL
      .....
      CPDGKCENTKPGSFKCIACQPGYRSQGGGACRDVNECAEGSPGWCENLPGSFRCCTCAQGYAPAPDGR-
      700      710      720      730      740      750      760

      770      780      790      800      810      820      830
inputs SCIDVDDCEAGKVCQDGICTNTPGSFQCQCLSGYHLRDRSRCEDIDECDFPAACIGGDCINTNGSYRCL
      .....
      SCLDVDECEAGDVCDNGICSNTPGSFQCQCLSGYHLRDRSHCEDIDECDFPAACIGGDCINTNGSYRCL
      770      780      790      800      810      820      830

      840      850      860      870      880      890
inputs CPLGHRVLVGGRKCKKIDIDECSDQDPLCLPH-ACENLQGSYVCVCDEGFTLTQDQHGCCEEVEQPHHKKECY
      .....
      CPQGHRLVGGRKCKQ-DIDECSDQDPSLCLPHGACKNLQGSYVCVCDEGFTPTQDQHGCCEEVEQPHHKKECY
      840      850      860      870      880      890      900

      900      910      920      930      940      950      960
inputs LNFDDTVFCDSVLATNVTQQECCCSLGAGWGDHCEIYPCPVYSSAEFHS LVPD GKRLHSGQQHCELCIPA
      .....
      LNFDDTVFCDSVLATNVTQQECCCSLGAGWGDHCEIYPCPVYSSAEFHS LCPDGKGYTQDNNIVNYGIPA
      910      920      930      940      950      960      970

      970      980      990      1000      1010      1020      1030
inputs HRDIDECILFGAEICKEGKCVNSQPGYECYCKQGFYYDGNLLECVDVDECLDESNCRNGVCENTWR-LPC
      .....
      HRDIDECMLFGSEICKEGKCVNTQPGYECYCKQGFYYDGNLLECVDVDECLDESNCRNGVCENTRGGYRC
      980      990      1000      1010      1020      1030      1040

      1040      1050      1060
inputs ACTPPAEYSPAQAQCLSPEEM-----EH
      .....
      ACTPPAEYSPAQRQCLSPEEMDVDECQDPAACRPGRCVNLPGSYRCECRPPWVPGPSGRDCQLPESPAER
      1050      1060      1070      1080      1090      1100      1110

      1070      1080      1090      1100      1110      1120      1130
inputs APERREVCWGQRGEDGMCMPLAGPALTFDDCCCRQRL-GYQCRPCPPRGTSQCPTSQSESNSFWDT
      .....
      APERRDVCWSQRGEDGMCAGPLAGPALTFDDCCCRQGRGWAQCRPCPPRGAGSHCPTSQSESNSFWDT
      1120      1130      1140      1150      1160      1170      1180

```

FIGURE 15B

	1140	1150	1160	1170	1180	1190	1200
inputs	PLLLGKSPRDEDSSEEDSDECRCVSGPCVPRPGGAVCECPGGFQLDASRARCVDIDECRELNQRGLLCKS ::: PLLLGKPPRDEDSSEEDSDECRCVSGRCVPRPGGAACECPGGFQLDASRARCVDIDECRELNQRGLLCKS 1190 1200 1210 1220 1230 1240 1250						
	1210	1220	1230	1240	1250		
inputs	ERCVNTSGSFRCVCKAGFTRSRPHGPACLSAAADDAAIAHTSVIDHRGYFH ::: ERCVNTSGSFRCVCKAGFARSRPHG-ACVP-----QRR---R 1260 1270 1280						

FIGURE 15C

mouse T275 seq

```

cctcctgctg tccccctcct acccttggt tctcgccccg ctctgccctc tgctaccaac
60
actcgatccc ctgctcgggc tcgacctcca atctccgagg gtcgtgcggc cccggatgcc
120
cgggccccga gcggtgcccc cggcctggcc cctgcg atg cgc cag gcc ggc gga
174
                                Met Arg Gln Ala Gly Gly
                                1         5

ttg ggg ctg ctg gca cta ctc ctg ctg gcg ctg ctg ggc ccc ggc ggc
222
Leu Gly Leu Leu Ala Leu Leu Leu Leu Ala Leu Leu Gly Pro Gly Gly
      10         15         20

cga ggg gtg ggc cgg ccg ggc agc ggg gca cag gcg ggg gcg ggg cgc
270
Arg Gly Val Gly Arg Pro Gly Ser Gly Ala Gln Ala Gly Ala Gly Arg
      25         30         35

tgg gcc caa cgc ttc aag gtg gtc ttt gcg cct gtg atc tgc aag cgg
318
Trp Ala Gln Arg Phe Lys Val Val Phe Ala Pro Val Ile Cys Lys Arg
      40         45         50

acc tgt ctg aag ggc cag tgt cgg gac agc tgt cag cag ggc tcc aac
366
Thr Cys Leu Lys Gly Gln Cys Arg Asp Ser Cys Gln Gln Gly Ser Asn
      55         60         65         70

atg acg ctc atc gga gag aac ggc cac agc acc gac acg ctc acc ggt
414
Met Thr Leu Ile Gly Glu Asn Gly His Ser Thr Asp Thr Leu Thr Gly
      75         80         85

tct gcc ttc cgc gtg gtg gtg tgc cct cta ccc tgc atg aac ggt ggc
462
Ser Ala Phe Arg Val Val Val Cys Pro Leu Pro Cys Met Asn Gly Gly
      90         95         100

cag tgc tct tcc cga aac cag tgc ctg tgt ccc ccg gat ttc acg ggg
510
Gln Cys Ser Ser Arg Asn Gln Cys Leu Cys Pro Pro Asp Phe Thr Gly
      105         110         115

cgc ttc tgc cag gtg cct gct gca gga acc gga gct ggc acc ggg agt
558
Arg Phe Cys Gln Val Pro Ala Ala Gly Thr Gly Ala Gly Thr Gly Ser
      120         125         130

tca ggc ccc ggc tgg ccc gac cgg gcc atg tcc aca ggc ccg ctg ccg
606
Ser Gly Pro Gly Trp Pro Asp Arg Ala Met Ser Thr Gly Pro Leu Pro
      135         140         145         150

ccc ctt gcc cca gaa gga gag tct gtg gct agc aaa cac gcc att tac
654
Pro Leu Ala Pro Glu Gly Glu Ser Val Ala Ser Lys His Ala Ile Tyr
      155         160         165

gcg gtg cag gtg atc gca gat cct ccc ggg ccg ggg gag ggt cct cct
702
Ala Val Gln Val Ile Ala Asp Pro Pro Gly Pro Gly Glu Gly Pro Pro
      170         175         180

```

FIGURE 16 A

gca caa cat gca gcc ttc ttg gtg ccc ctg ggg cca gga caa atc tcg
 750
 Ala Gln His Ala Ala Phe Leu Val Pro Leu Gly Pro Gly Gln Ile Ser
 185 190 195

gca gaa gtg cag gct ccg ccc ccc gtg gtg aac gtg cgt gtc cat cac
 798
 Ala Glu Val Gln Ala Pro Pro Pro Val Val Asn Val Arg Val His His
 200 205 210

cct cct gaa gct tcc gtt cag gtg cac cgc atc gag ggg ccg aac gct
 846
 Pro Pro Glu Ala Ser Val Gln Val His Arg Ile Glu Gly Pro Asn Ala
 215 220 225 230

gaa ggc cca gcc tct tcc cag cac ttg ctg ccg cat ccc aag ccc ccg
 894
 Glu Gly Pro Ala Ser Ser Gln His Leu Leu Pro His Pro Lys Pro Pro
 235 240 245

cac ccg agg cca ccc act caa aag cca ctg ggc cgc tgc ttc cag gac
 942
 His Pro Arg Pro Pro Thr Gln Lys Pro Leu Gly Arg Cys Phe Gln Asp
 250 255 260

aca ttg ccc aag cag cct tgt ggc agc aac cct ttg cct ggc ctt acc
 990
 Thr Leu Pro Lys Gln Pro Cys Gly Ser Asn Pro Leu Pro Gly Leu Thr
 265 270 275

aag cag gaa gat tgc tgc ggt agc atc ggt act gcc tgg gga caa agc
 1038
 Lys Gln Glu Asp Cys Cys Gly Ser Ile Gly Thr Ala Trp Gly Gln Ser
 280 285 290

aag tgt cac aag tgc cca cag ctt cag tat aca ggg gtg cag aag cct
 1086
 Lys Cys His Lys Cys Pro Gln Leu Gln Tyr Thr Gly Val Gln Lys Pro
 295 300 305 310

gta cct gta cgt ggg gag gtg ggt gct gac tgc ccc cag ggc tac aag
 1134
 Val Pro Val Arg Gly Glu Val Gly Ala Asp Cys Pro Gln Gly Tyr Lys
 315 320 325

agg ctc aac agc acc cac tgc cag gat atc aac gaa tgt gcg atg ccc
 1182
 Arg Leu Asn Ser Thr His Cys Gln Asp Ile Asn Glu Cys Ala Met Pro
 330 335 340

ggg aat gtg tgc cat ggt gac tgc ctc aac aac cct ggc tct tat cgc
 1230
 Gly Asn Val Cys His Gly Asp Cys Leu Asn Asn Pro Gly Ser Tyr Arg
 345 350 355

tgt gtc tgc ccg ccc ggt cat agc ttg ggt ccc ctc gca gca cag tgc
 1278
 Cys Val Cys Pro Pro Gly His Ser Leu Gly Pro Leu Ala Ala Gln Cys
 360 365 370

att gcc gac aaa cca gag gag aag agc ctg tgt ttc cgc ctt gtg agc
 1326
 Ile Ala Asp Lys Pro Glu Glu Lys Ser Leu Cys Phe Arg Leu Val Ser
 375 380 385 390

FIGURE 16B

acc gaa cac cag tgc cag cac cct ctg acc aca cgc cta acc cgc cag
 1374 Thr Glu His Gln Cys Gln His Pro Leu Thr Thr Arg Leu Thr Arg Gln
 395 400 405
 ctc tgc tgc tgt agt gtg ggt aaa gcc tgg ggt gcc cgg tgc cag cgc
 1422 Leu Cys Cys Cys Ser Val Gly Lys Ala Trp Gly Ala Arg Cys Gln Arg
 410 415 420
 tgc ccg gca gat ggt aca gca gcc ttc aag gag atc tgc ccc ggc tgg
 1470 Cys Pro Ala Asp Gly Thr Ala Ala Phe Lys Glu Ile Cys Pro Gly Trp
 425 430 435
 gaa agg gta cca tat cct cac ctc cca cca gac gct cac cat cca ggg
 1518 Glu Arg Val Pro Tyr Pro His Leu Pro Pro Asp Ala His His Pro Gly
 440 445 450
 gga aag cga ctt ctc cct ctt cct gca ccc gac ggg cca ccc aaa ccc
 1566 Gly Lys Arg Leu Leu Pro Leu Pro Ala Pro Asp Gly Pro Pro Lys Pro
 455 460 465 470
 cag cag ctt cct gaa agc ccc agc cga gca cca ccc ctc gag gac aca
 1614 Gln Gln Leu Pro Glu Ser Pro Ser Arg Ala Pro Pro Leu Glu Asp Thr
 475 480 485
 gag gaa gag aga gga gtg acc atg gat cca cca gtg agt gag gag cga
 1662 Glu Glu Glu Arg Gly Val Thr Met Asp Pro Pro Val Ser Glu Glu Arg
 490 495 500
 tcg gtg cag cag agc cac ccc act acc acc acc tca ccc ccc cgg cct
 1710 Ser Val Gln Gln Ser His Pro Thr Thr Thr Thr Ser Pro Pro Arg Pro
 505 510 515
 tac cca gag ctc atc tct cgc ccc tcc cca cct acc ttc cac cgg ttc
 1758 Tyr Pro Glu Leu Ile Ser Arg Pro Ser Pro Pro Thr Phe His Arg Phe
 520 525 530
 ctg cca gac ttg ccc cca tcc cga agt gca gtg gag atc gcc ccc act
 1806 Leu Pro Asp Leu Pro Pro Ser Arg Ser Ala Val Glu Ile Ala Pro Thr
 535 540 545 550
 cag gtc aca gag acc gat gag tgc cga ttg aac cag aat atc tgt ggc
 1854 Gln Val Thr Glu Thr Asp Glu Cys Arg Leu Asn Gln Asn Ile Cys Gly
 555 560 565
 cat gga cag tgt gtg cct ggc ccc tcg gat tac tcc tgc cac tgc aac
 1902 His Gly Gln Cys Val Pro Gly Pro Ser Asp Tyr Ser Cys His Cys Asn
 570 575 580
 gct ggc tac cgg tca cac ccg cag cac cgc tac tgt gtt gat gtg aac
 1950 Ala Gly Tyr Arg Ser His Pro Gln His Arg Tyr Cys Val Asp Val Asn
 585 590 595

FIGURE 16C

gag tgc gag gca gag ccc tgc ggc ccc ggg aaa ggc atc tgt atg aac
 1998
 Glu Cys Glu Ala Glu Pro Cys Gly Pro Gly Lys Gly Ile Cys Met Asn
 600 605 610

act ggt ggc tcc tac aat tgt cac tgc aac cga ggc tac cgc ctc cac
 2046
 Thr Gly Gly Ser Tyr Asn Cys His Cys Asn Arg Gly Tyr Arg Leu His
 615 620 625 630

gtg ggt gca ggg ggc cgc tcg tgc gtg gac ctg aac gag tgc gcc aag
 2094
 Val Gly Ala Gly Gly Arg Ser Cys Val Asp Leu Asn Glu Cys Ala Lys
 635 640 645

cct cac ctg tgt ggg gac ggt ggc ttc tgc atc aac ttc cct ggt cac
 2142
 Pro His Leu Cys Gly Asp Gly Gly Phe Cys Ile Asn Phe Pro Gly His
 650 655 660

tac aaa tgc aac tgc tat cct ggc tac cgg ctc aag gcc tcc cga ccg
 2190
 Tyr Lys Cys Asn Cys Tyr Pro Gly Tyr Arg Leu Lys Ala Ser Arg Pro
 665 670 675

ccc att tgc gaa gac atc gac gag tgt cgc gac cct agc acc tgc cct
 2238
 Pro Ile Cys Glu Asp Ile Asp Glu Cys Arg Asp Pro Ser Thr Cys Pro
 680 685 690

gat ggc aaa tgt gaa aac aaa cct ggc agc ttc aag tgc atc gcc tgc
 2286
 Asp Gly Lys Cys Glu Asn Lys Pro Gly Ser Phe Lys Cys Ile Ala Cys
 695 700 705 710

cag cct ggc tac cgt agc cag ggg ggc ggg gcc tgt cgt gat gtc aac
 2334
 Gln Pro Gly Tyr Arg Ser Gln Gly Gly Gly Ala Cys Arg Asp Val Asn
 715 720 725

gaa tgc tcc gag ggt acc ccc tgc tct cct gga tgg tgt gag aac ctt
 2382
 Glu Cys Ser Glu Gly Thr Pro Cys Ser Pro Gly Trp Cys Glu Asn Leu
 730 735 740

ccg ggt tct tac cgt tgc acg tgt gcc cag ggg ata cga acc cgc aca
 2430
 Pro Gly Ser Tyr Arg Cys Thr Cys Ala Gln Gly Ile Arg Thr Arg Thr
 745 750 755

gga cgc ctc agt tgc ata gac gtg gat gag tgt gag gct ggg aaa gtg
 2478
 Gly Arg Leu Ser Cys Ile Asp Val Asp Glu Cys Glu Ala Gly Lys Val
 760 765 770

tgc caa gat ggc atc tgc acg aac aca cca ggc tct ttc cag tgt cag
 2526
 Cys Gln Asp Gly Ile Cys Thr Asn Thr Pro Gly Ser Phe Gln Cys Gln
 775 780 785 790

tgc ctc tcc ggc tat cat ctg tca agg gat cgg agc cgc tgt gag gac
 2574
 Cys Leu Ser Gly Tyr His Leu Ser Arg Asp Arg Ser Arg Cys Glu Asp
 795 800 805

FIGURE 16D

```

att gat gaa tgt gac ttc cct gcg gcc tgc atc ggg ggt gac tgc atc
2622
Ile Asp Glu Cys Asp Phe Pro Ala Ala Cys Ile Gly Gly Asp Cys Ile
      810      815      820

aat acc aat ggt tcc tac aga tgt ctc tgt ccc ctg ggt cat cgg ttg
2670
Asn Thr Asn Gly Ser Tyr Arg Cys Leu Cys Pro Leu Gly His Arg Leu
      825      830      835

gtg ggc ggc agg aag tgc aag aaa gat ata gat gag tgc agc cag gac
2718
Val Gly Gly Arg Lys Cys Lys Lys Asp Ile Asp Glu Cys Ser Gln Asp
      840      845      850

cca ggc ctg tgc ctg ccc cat gcc tgc gag aac ctc cag ggc tcc tat
2766
Pro Gly Leu Cys Leu Pro His Ala Cys Glu Asn Leu Gln Gly Ser Tyr
      855      860      865      870

gtc tgt gtc tgt gat gag ggt ttc aca ctc acc cag gac cag cat ggg
2814
Val Cys Val Cys Asp Glu Gly Phe Thr Leu Thr Gln Asp Gln His Gly
      875      880      885

tgt gag gag gtg gag cag ccc cac cac aag aag gag tgc tac ctt aac
2862
Cys Glu Glu Val Glu Gln Pro His His Lys Lys Glu Cys Tyr Leu Asn
      890      895      900

ttc gat gac aca gtg ttc tgt gac agc gta ttg gct acc aat gtc act
2910
Phe Asp Asp Thr Val Phe Cys Asp Ser Val Leu Ala Thr Asn Val Thr
      905      910      915

cag cag gaa tgc tgt tgc tct ctg gga gct ggc tgg gga gac cac tgc
2958
Gln Gln Glu Cys Cys Cys Ser Leu Gly Ala Gly Trp Gly Asp His Cys
      920      925      930

gaa atc tat ccc tgt cca gtc tac agc tca gcc gaa ttt cac agc ctg
3006
Glu Ile Tyr Pro Cys Pro Val Tyr Ser Ser Ala Glu Phe His Ser Leu
      935      940      945      950

gtg cct gat ggg aaa agg cta cac tca gga caa caa cat tgt gaa cta
3054
Val Pro Asp Gly Lys Arg Leu His Ser Gly Gln Gln His Cys Glu Leu
      955      960      965

tgc att cct gcc cac cgt gac atc gac gaa tgc ata ttg ttt ggg gca
3102
Cys Ile Pro Ala His Arg Asp Ile Asp Glu Cys Ile Leu Phe Gly Ala
      970      975      980

gag atc tgc aag gag ggc aag tgt gtg aac acg cag ccc ggc tac gag
3150
Glu Ile Cys Lys Glu Gly Lys Cys Val Asn Thr Gln Pro Gly Tyr Glu
      985      990      995

tgc tac tgc aag cag ggc ttc tac tac gat ggc aac ctg ctg gag tgc
3198
Cys Tyr Cys Lys Gln Gly Phe Tyr Tyr Asp Gly Asn Leu Leu Glu Cys
      1000      1005      1010

```

FIGURE 1GE

gtg gac gtg gat gag tgc ttg gat gag tct aac tgc agg aac gga gtg
 3246
 Val Asp Val Asp Glu Cys Leu Asp Glu Ser Asn Cys Arg Asn Gly Val
 015 1020 1025 1030

tgt gag aac aca cgt ggc ggc tac cgc tgt gcc tgc act ccg ccg gca
 3294
 Cys Glu Asn Thr Arg Gly Gly Tyr Arg Cys Ala Cys Thr Pro Pro Ala
 1035 1040 1045

gag tac agt ccc gca cag gcc cag tgt ctg atc ccg gag aga tgg agc
 3342
 Glu Tyr Ser Pro Ala Gln Ala Gln Cys Leu Ile Pro Glu Arg Trp Ser
 1050 1055 1060

acg ccc cag aga gac gtg aag tgt gct ggg gcc agc gag gag agg acg
 3390
 Thr Pro Gln Arg Asp Val Lys Cys Ala Gly Ala Ser Glu Glu Arg Thr
 1065 1070 1075

gca tgt gta tgg ggc ccc tgg gcg gga cct gcc ctc act ttt gat gac
 3438
 Ala Cys Val Trp Gly Pro Trp Ala Gly Pro Ala Leu Thr Phe Asp Asp
 1080 1085 1090

tgc tgc tgc cgc cag ccg cgg ctg ggt acc cag tgc aga ccg tgc ccg
 3486
 Cys Cys Cys Arg Gln Pro Arg Leu Gly Thr Gln Cys Arg Pro Cys Pro
 095 1100 1105 1110

cca cgt ggc acc ggg tcc cag tgc ccg act tca cag agt gag agc aat
 3534
 Pro Arg Gly Thr Gly Ser Gln Cys Pro Thr Ser Gln Ser Glu Ser Asn
 1115 1120 1125

tct ttc tgg gac aca agc ccc ctg cta ctg ggg aag tct ccg cga gac
 3582
 Ser Phe Trp Asp Thr Ser Pro Leu Leu Leu Gly Lys Ser Pro Arg Asp
 1130 1135 1140

gaa gac agc tca gag gag gat tca gat gag tgc cgt tgt gtg agc gga
 3630
 Glu Asp Ser Ser Glu Glu Asp Ser Asp Glu Cys Arg Cys Val Ser Gly
 1145 1150 1155

cgc tgt gtg cca cgg cca ggc ggg gcg gta tgc gag tgt cct gga ggc
 3678
 Arg Cys Val Pro Arg Pro Gly Gly Ala Val Cys Glu Cys Pro Gly Gly
 1160 1165 1170

ttt cag ctg gac gcc tcc cgt gcc cgc tgc gtg gac att gat gag tgc
 3726
 Phe Gln Leu Asp Ala Ser Arg Ala Arg Cys Val Asp Ile Asp Glu Cys
 175 1180 1185 1190

cga gaa ctg aac cag cgg gga ctg ctg tgt aag agc gag cgg tgc gtg
 3774
 Arg Glu Leu Asn Gln Arg Gly Leu Leu Cys Lys Ser Glu Arg Cys Val
 1195 1200 1205

aac acc agt gga tcc ttc cgc tgt gtc tgc aaa gct ggc ttc acg cgc
 3822
 Asn Thr Ser Gly Ser Phe Arg Cys Val Cys Lys Ala Gly Phe Thr Arg
 1210 1215 1220

FIGURE 16F

agc cgc cct cac ggg cct gcg tgc ctc agc gcc gcc gct gat gat gca
3870
Ser Arg Pro His Gly Pro Ala Cys Leu Ser Ala Ala Ala Asp Asp Ala
1225 1230 1235
gcc ata gcc cac acc tca gtg atc gat cat cga ggg tat ttt cac
3915
Ala Ile Ala His Thr Ser Val Ile Asp His Arg Gly Tyr Phe His
1240 1245 1250
tgaaagtgga gacagacaag tacatccttt gctcctgacc aaacgagagc atggacccaa
3975
ggatccttca gggcccacaa atctccttcc cacaccccaa acccaagggtg ctctgtctg
4035
cagagtgtg tctgctttct cccaagggtg attcctagaa acttcgacat cagatctgcc
4095
cctttaattt actcttggct ttcaaggcaa attgatattc acatccaaag cgggcagcat
4155
caactgcttg gcgggttga ctgagctggg acccaggatg tgaaatagaa tttattgtgg
4215
ctctgattat gtacactaga tgtgcctgac ctgctgacca ggctcacatg gtttgtacaa
4275
taaatacatc cgccgggaaa aaaaaaaaaa aaaaaaaaaa a
1

FIGURE 16G

[illegible]

37/60

S	T	P	P	P	P	S	S	R	*	362
TCC	ACA	CCG	CCA	CCA	CCA	AGC	AGC	CGC	TGA	1116
GACGGACGGTTCCATGCCAGCTGCCCTGGAGGAGGAACAGACCCCTTTAGTCCTCATCCCTTAGATCCTGGAGGGCACGG										1195
ATCACATCCTGGGAAGAAGGCATCTGGAGGATAAGCAAAGCCACCCCGACACCCAATCTTGGAAGCCCTGAGTAGGCAG										1274
GGCCAGGGTAGGTGGGGGCCGGGAGGGACCCAGGTGTGAACGGATGAATAAAGTTCAA										1332

FIGURE 17B

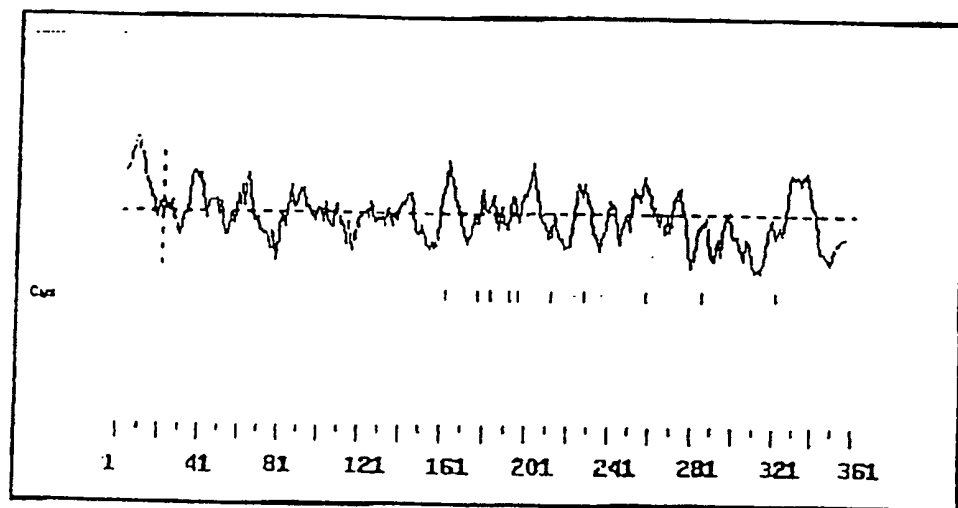


FIGURE 18

mouse TANGO 300 sequence

```

gtcgacccac gcgtccgcat ccaccagcag aaatcctgtc atg gcg aga ctc ggg      55
                                     Met Ala Arg Leu Gly
                                     1       5

.ctg ctt ctc ctc ctg ctg ctg gcc ctg cca cca cac ttc tcc tca gtg
103
Leu Leu Leu Leu Leu Leu Ala Leu Pro Pro His Phe Ser Ser Val
      10              15              20

tca tgg cca gac act gca cag ggc acc atg gca aac ttg atc ctg act
151
Ser Trp Pro Asp Thr Ala Gln Gly Thr Met Ala Asn Leu Ile Leu Thr
      25              30              35

gca tta gaa aaa gcc acc ttg ttc ttg gag gac agg ctg ccc aca atc
199
Ala Leu Glu Lys Ala Thr Leu Phe Leu Glu Asp Arg Leu Pro Thr Ile
      40              45              50

aac ctg gat ggt gtg gtg ggc ttc caa gtg ctg gaa gtg caa ctc cga
247
Asn Leu Asp Gly Val Val Gly Phe Gln Val Leu Glu Val Gln Leu Arg
      55              60              65

gga gtt cag gaa aaa tgg gct cac aag ccc ttg ctg cag cct ctc agc
295
Gly Val Gln Glu Lys Trp Ala His Lys Pro Leu Leu Gln Pro Leu Ser
      70              75              80              85

atg cgc gct gga cag atg gcc aac aca ctg tct gct ctc ctc caa aaa
343
Met Arg Ala Gly Gln Met Ala Asn Thr Leu Ser Ala Leu Leu Gln Lys
      90              95              100

tcc atc ttc tac ctc aag cag agt gac ccc acg tac cta aga gag ttc
391
Ser Ile Phe Tyr Leu Lys Gln Ser Asp Pro Thr Tyr Leu Arg Glu Phe
      105              110              115

cag cca agc att cag cct ggg ttt tgg aag ttg ccc aat gac tgg aca
439
Gln Pro Ser Ile Gln Pro Gly Phe Trp Lys Leu Pro Asn Asp Trp Thr
      120              125              130

cgc ace aat gcc tcc cta gtc tac ccc tgg ctg gaa ccc ctg gac tct
487
Arg Thr Asn Ala Ser Leu Val Tyr Pro Trp Leu Glu Pro Leu Asp Ser
      135              140              145

ttc tca gag gaa agc agc gat gtg tgc ctg gtg caa cta cta gga aca
535
Phe Ser Glu Glu Ser Ser Asp Val Cys Leu Val Gln Leu Leu Gly Thr
      150              155              160              165

ggg aca gac agc agc cag cct tgc agg ctc tcc aac ttc tgc aga acc
583
Gly Thr Asp Ser Ser Gln Pro Cys Arg Leu Ser Asn Phe Cys Arg Thr
      170              175              180

ctt atg acc aag gcc ggc tgc tca ggc tac agc ctc tcc cat cag ctg
631
Leu Met Thr Lys Ala Gly Cys Ser Gly Tyr Ser Leu Ser His Gln Leu
      185              190              195

```

FIGURE 19A

ctc ttc ttc ctc tgg gcc aga atg caa ggg tgc acg gag gga ctg ttc
 679
 Leu Phe Phe Leu Trp Ala Arg Met Gln Gly Cys Thr Glu Gly Leu Phe
 200 205 210

ctc cag agc caa cac tac atg gac atc ttc tgt gcc aat atg atg gaa
 727
 Leu Gln Ser Gln His Tyr Met Asp Ile Phe Cys Ala Asn Met Met Glu
 215 220 225

ctg aac cac aga gct gag gcc gtt gga tac gct tac ccc acc caa gac
 775
 Leu Asn His Arg Ala Glu Ala Val Gly Tyr Ala Tyr Pro Thr Gln Asp
 230 235 240 245

ctc ttc atg gaa aac att atg ttc tgt ggt atg gct ggc ttc tct gac
 823
 Leu Phe Met Glu Asn Ile Met Phe Cys Gly Met Ala Gly Phe Ser Asp
 250 255 260

ttc tac aag ctg cgc tgg ctg gag gcc att ctc agc tgg cag aac ccc
 871
 Phe Tyr Lys Leu Arg Trp Leu Glu Ala Ile Leu Ser Trp Gln Asn Pro
 265 270 275

cag gtg gga tgc ttc ggg agg cct gac aca aag ggt gaa cct tct gaa
 919
 Gln Val Gly Cys Phe Gly Arg Pro Asp Thr Lys Gly Glu Pro Ser Glu
 280 285 290

gtt cca cat cag cag ggc att ctg aga aga gtg cga agg cgg gaa aaa
 967
 Val Pro His Gln Gln Gly Ile Leu Arg Arg Val Arg Arg Arg Glu Lys
 295 300 305

ctg ttc gca gat ggc tgt tcg tgc cac aac aca gcc aca gca gtc gca
 1015
 Leu Phe Ala Asp Gly Cys Ser Cys His Asn Thr Ala Thr Ala Val Ala
 310 315 320 325

gcc ctg ggt ggc ttt ctc tac atc ctg gca gaa tac cac cca gac aat
 1063
 Ala Leu Gly Gly Phe Leu Tyr Ile Leu Ala Glu Tyr His Pro Asp Asn
 330 335 340

gga gat gca cat cca gaa tac tac cca aac cat gga gat cca tac tca
 1111
 Gly Asp Ala His Pro Glu Tyr Tyr Pro Asn His Gly Asp Pro Tyr Ser
 345 350 355

tcc tca cag tca cca gca agc aac tac caa gat ggt gct gcc ggc cct
 1159
 Ser Ser Gln Ser Pro Ala Ser Asn Tyr Gln Asp Gly Ala Ala Gly Pro
 360 365 370

gac gtc cag agg act ggc agg ccc ctt agt gtt tct taagtccctga
 1205
 Asp Val Gln Arg Thr Gly Arg Pro Leu Ser Val Ser
 375 380 385

gtcagaggtc acaggctgag gaggcaattg aggaaagtga ccagctatat ccccatcgcc
 1265
 acttctgggt gtttaaaagt cttgggagag cagggccagg gaaagcaggg ttggagagtg
 1325

FIGURE 19B

gggtggccca gatgtcagca gaatacataa agcacagtca attggagctg aaaaaaaaaa
1385
aaaaaggcg gccgc
1400

FIGURE 19C

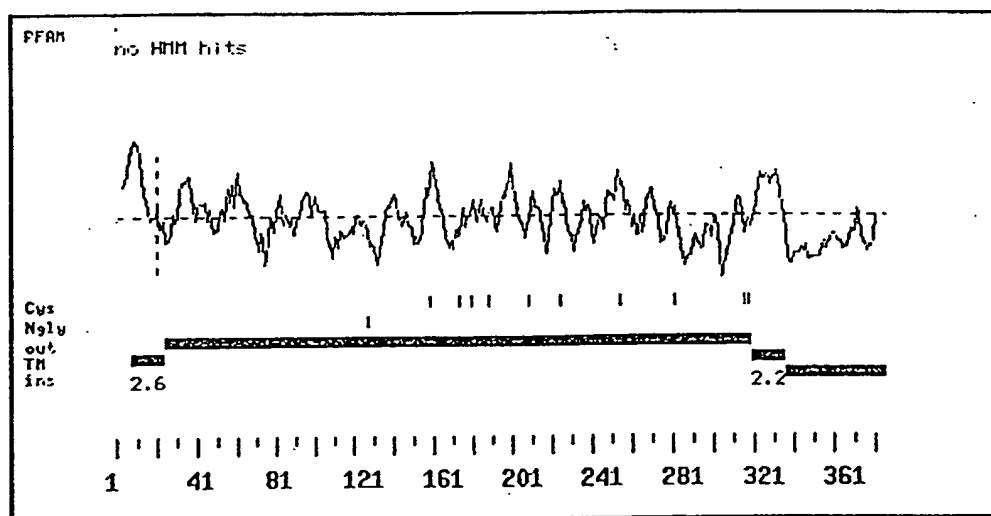


FIGURE 20

mT300 ORF (analysis only) - Import - complete

Gap Weight:	12	Average Match:	10.000
Length Weight:	4	Average Mismatch:	0.000
Quality:	8110	Length:	1174
Ratio:	7.488	Gaps:	7
Percent Similarity:	77.726	Percent Identity:	77.726

[illegible]

FIGURE 21 A

439 CTGGACTCTTTCTCAGAGGAAAGCAGCGATGTGTGCCTGGTGCAACTACT 488
501 GGGAACCGGGACGGACAGCAGCGAGCCCTGCGGCCTCTCAGACCTCTGCA 550
489 AGGAACAGGGACAGACAGCAGCCAGCCTTGCAAGGCTCTCCAATTCTGCA 538
551 GGAGCCTCATGACCAAGCCCGGCTGCTCAGGCTACTGCCTGTCCCACCAA 600
539 GAACCCCTTATGACCAAGCCCGGCTGCTCAGGCTACAGCCTCTCCCATCAG 588
601 CTGCTCTTCTTCTCTGCGCCAGAATGAGGGGGTGACACAGGGACCCT 650
589 CTGCTCTTCTTCTCTGCGCCAGAATGCAAGGGTGACAGGAGGGACTGTT 638
651 CCAACAGAGCCAGGACTATATCAACCTCTTCTGCGCCAACATGATGGACT 700
639 CCTCCAGAGCCAACACTACATGGACATCTTCTGTGCCAATATGATGGAAC 688
701 TGAACCGCAGAGCTGAGGCCATCGGATACGCCTACCCTACCCGGGACATC 750
689 TGAACCACAGAGCTGAGGCCGTTGGATACGCTTACCCACCCAAGACCTC 738
751 TTCATGGAACACATCATGTTCTGTGGAATGGGCGGCTTCTCCGACTTCTA 800
739 TTCATGGAACACATTATGTTCTGTGGTATGGCTGGCTTCTCTGACTTCTA 788
801 CAAGCTCCGGTGGCTGGAGGCCATTCTCAGCTGGCAGAAACAGCAGGAAG 850
789 CAAGCTGCGCTGGCTGGAGGCCATTCTCAGCTGGCAGAACCCCAAGGTGG 838
851 GATGCTTCGGGGAGCCTGATGCTGAAGATGAAGAATTATCTAAAGCTATT 900
839 GATGCTTCGGGGAGGCCTGACAC...AAAGGGTGAACCTTCTGAAG...TT 882
901 CAATATCAGCAGCATTTTTTCGAGGAGAGTGAAGAGGCGAGAAAAACAATT 950
883 CCACATCAGCAGGGCATTCTGAGAAGAGTGCGAAGGCGGGAAAAACTGTT 932
951 TCCAGATGGCTGCTCCTCCCAACAACAGCCACAGCAGTGGCAGCCCTGG 1000
933 CGCAGATGGCTGTTCTGTCACAACACAGCCACAGCAGTCGCAGCCCTGG 982
1001 GTGGCTTCTTATACATCCTGGCAGAATACCCCCAGCAAACAGAGAGCCA 1050
983 GTGGCTTCTTCTACATCCTGGCAGAATACCACCCAGACAATGGAGATGCA 1032
1051 CACCCATCCACACCGCCACCACCAAGCAGCCGC..... 1083
1033 CATCCA.GAATACTACCCAAACCATGGAGATCCATACTCATCCTCACAGT 1081

FIGURE 21.B

GAP of: FrGcgManager_687_IBFGliaq_ check: 8297 from: 1 to: 361

ht300 prot (analysis only) - Import - complete

to: FrGcgManager_687_JBFmT7mm_ check: 9127 from: 1 to: 385

mT300 prot (analysis only) - Import - complete

Gap Weight:	12	Average Match:	2.778
Length Weight:	4	Average Mismatch:	-2.248
Quality:	1237	Length:	391
Ratio:	3.427	Gaps:	4
Percent Similarity:	74.930	Percent Identity:	69.577

```

HUMAN 1 MASLGLLLLLLLLTALPPLWSSSLPGLDTAESKATIADLILSALERATVFL 50
      || ||||| ||||| . ||| . |||: |.|. |||. |||: ||. ||
MOUSE 1 MARLGLLLLLLL .ALPPHF.SSVSWPDTAQ..GTMANLILTALEKATLFL 46

51 EQRLPEINLDGMVGVRVLEEQLKSVREKWAQEP LLOPLSLRVGMLGEKLE 100
      || ||| ||||| . ||| |||: |. |||| . |||||: | | : |
47 EDRLPTINLDGVVGFQVLEVLQVLRGVQEKWAHKPLLQPLSMRAGQMANTLS 96

101 AAIQ RSLHYLKLSDPKYLREFQLTLQPGFWKLP HAWIHTDASLVYPTFGP 150
      | :|:|: || ||| ||||| .: ||||| . | |. |||||
97 ALLQKSI FYLKQSDPTYLREFQPSIQPGFWKLPNDWTRTNASLVYPWLEP 146

151 QDSFSEERSDVCLVQLLGTGTDSSSEPCGLSDLCRSLMTKPGCSGYCLSHQ 200
      ||||| ||||| ||||| |||||: || ||. ||. |||| ||||| |||||
147 LDSFSEESSDVCLVQLLGTGTDSSQPCRLSNFCRTLMTKAGCSGYSLSHQ 196

201 LLFFLWARMRGCTQGPLQQSQDYINLFCANMMDLNRRAEAIGYAYPTRDI 250
      ||||| ||||| . |||: | ||| |..: |||||: || ||||: ||||| .|:
197 LLFFLWARMQGCTEGLFLQSQHYMDIFCANMELNHRAEAVGYAYPTQDL 246

251 FMENIMFCGMGGFSDFYKLRWLEAILSWQKQEGCFGE PD AEDEELSKAI 300
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
247 FMENIMFCGMAGFSDFYKLRWLEAILSWQNPQVGCFGRPD TKGE..PSEV 294

301 QYQQHFSRRVKRREKQFPDGCSSHNTATAVAALGGFLYILA EYPPANREP 350
      :|| |||: ||||| ||||| ||||| ||||| ||||| ||||| :
295 PHQQGILRRVRRREKLFADGCSCHNTATAVAALGGFLYILA EYHPDNGDA 344

351 HPSTPPPPSSR..... 361
      || |
345 HPEYYPNHGDPYSSSQSPASNYQDGAAGPDVQRTGRPLSVS 385
  
```

FIGURE 22

Input file T245Alhbab165e5; Output File T245Alhbab165e5.pac
Sequence length 1356

GTCGACCCACGCGTCCGGCAGCCTGCAGCCCGCAGCCCGCAGCCCGGAGCCAGATCGCGGGGCTCAGACCAAAACCCGACT 79

CGAECGCGCGCCCCAGCCAGGCGCC ATG CTG CCG CTT CTG CTG GGC CTG CTG GGC CCA GCG GCC 143
M L P L L L G L L S P A A
C W A L G P T P G P G S S E L R E A F S
TGC TGG GCC CTG GGC CCG ACC CCC GGC CCG GGA TCC TCT GAG CTG CGC TCG GCC TTC TCG 243
A A R T T P L E G T S E M A V T F D K V
GCG GCA CGC ACC ACC CCC CTG GAG GGC ACG TCG GAG ATG GCG GTG ACC TTC GAC AAG GTG 243
Y V N I G G D F D V A T G Q F R C R V P
TAC GTG AAC ATC GGG GGC GAC TTC GAT GTG GCC ACC GGC CAG TTT CGC TGC CGC GTG CCC 323
G A Y F F S F T A G K A P H K S L S V M
GGC GCC TAC TTC TTC TCC TTC ACG GCT GGC AAG GCC CCG CAC AAG AGC CTG TCG GTG ATG 383
L V R N R D E V Q A L A F D E Q F R P G
CTG GTG CGA AAC CGC GAC GAG GTG CAG GCG CTG GCC TTC GAC GAG CAG CGG CGG CCA GGC 443
A R R A A S Q S A M L Q L D Y G D T V W
CGC CGG CGC GCA GCC AGC CAG AGC GCC ATG CTG CAG CTC GAC TAC GGC GAC ACA GTG TGG 503
L R L H G A P Q Y A L G A P G A T F S G
CTG CGG CTG CAT GGC GCC CCG CAG TAC GCG CTA GGC GCG CCC GGC GCC ACC TTC AGC GGC 563
Y L V Y A D A D A D A P A R G P P A P P
TAC CTA GTC TAC GCC GAC GCC GAC GCT GAC GCG CCT GCG CGC GGG CCG CCC GCG CCC CCC 623
E P R S A F S A A R T R S L V G S D A G
GAG CCG CGC TCG GCC TTC TCG GCG GCG CGC ACG CGC AGC TTG GTG GGC TCG GAC GCT GGC 683
P G P R H Q P L A F D T E F V N I G G D
CCC GGG CCG CGG CAC CAA CCA CTC GCC TTC GAC ACC GAG TTC GTC AAC ATT GGC GGC GAC 743
F D A A A G V F R C R L P G A Y F F S F
TTC GAC GCG GCG GCC GGC GTG TTC CGC TGC CGT CTG CCC GGC GCC TAC TTC TTC TCC TTC 803
T L G K L P R K T L S V K L M K N R D E
ACG CTG GGC AAG CTG CCG CGT AAG ACG CTG TCG GTT AAG CTG ATG AAG AAC CGC GAC GAG 863
V -Q A M I Y D D G A S R R R E M Q S Q S
GTG CAG GCC ATG ATT TAC GAC GAC GGC GCG TCG CGG CGC CGC GAG ATG CAG AGC CAG AGC 923
V M L A L R R G D A V W L L S H D H D G
GTG ATG CTG GCC CTG CGG CGC GGC GAC GCC GTC TGG CTG CTC AGC CAC GAC CAC GAC GGC 983
Y G A Y S N H D L P T D L K T V L P S W
TAC GGC GCC TAC AGC AAC CAC GAT CTC CCA ACT GAC CTC AAA ACG GTT TTG CCG AGT TGG 1043
D V H C C Q V N Q R F E L C I G V I P E
GAC GTC CAC TGC TGT CAA GTC AAC CAG AGA TTT GAA CTG TGC ATT GGT GTG ATC CCT GAG 1103
E S Q H W D D A I R M D T D L
GAA AGT CAG CAC TGG GAT GAC GCC ATC AGG ATG GAT ACA GAC CTC TAA 1163

CTCATTTGAAGCAGGACACCTGCACACATGAAAGTGAGGGGAGAGGGAACAAAGAGCTACTGAGGGAACAGCTAACTTCA 1230
GCTGGAGTCACCTGGTTTAATGCTGAGAGAAAAGTCCAAGCTTGGGATGGAGGAATCTGTAGTTTCTTTGAAACAAGTC 1309
TGCCCACTCCACAAAAAAAAAAAAAAAAAAAAAAAAAAGGGCGGCCGC 1356

FIGURE 23B

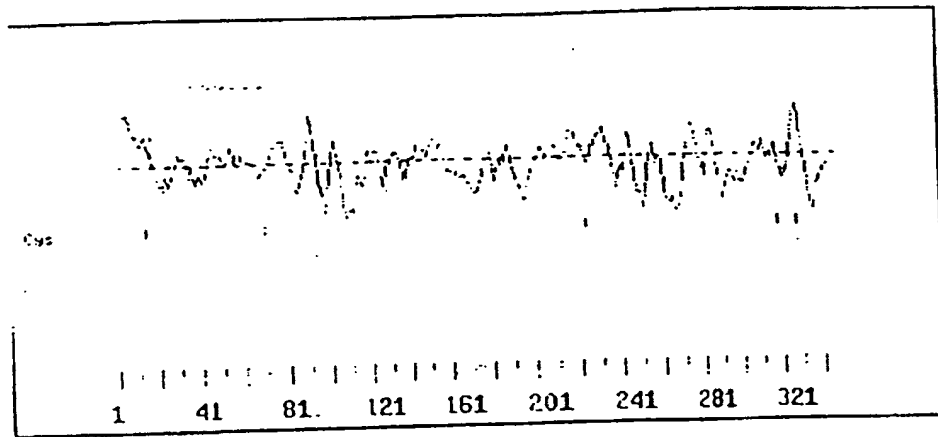


FIGURE 24

Input file T245A1kbd75hi; Output file T245A1kbd75hi.pac
Sequence length 1388

JTCGACCCACGCGTCCGGGCTCCCTCCCTGGGTGACCTTCCCTTCCCTCACGCCTCCCCACGCCCGGGCTGGGGCCCC 79
 AGCACCCCTGTCCGCCCGCCCTCAGAGCCGGGAGAGCAGCCGGAGCCCCCGGCGCCCCCGCCGAGCCGGGCGGTCA 158
 GTGCGCAGCCCGGCGCCCGCAGCCCGCAGCCCGCAGCCGGATCGCGGGCTCGGACCGAACCAGCCGACCGCCGCCCC 237
 M L P L L L G L L G P A A C H A
 CAGCCAGGCTCC ATG CTG CCG CTT CTG CTA GGC CTG CTG GGC CCA GCG GCC TGC TGG GCC 297
 L G P A P G P G S S E L R S A F S A A R
 CTG GGC CCG GCC CCC GGC CCG GGC TCC TCA GAG CTG CGC TCG GCC TTC TCG GCG GCA CGC 357
 T T P L E G A S E M A V T F D K V Y V N
 ACC ACT CCC CTG GAG GGC GCG TCG GAG ATG GCG GTG ACC TTC GAC AAG GTG TAC GTG AAC 417
 I G G D F D A A T G Q F R C R V P G A Y
 ATC GGG GGC GAC TTC GAC GCG GCC ACA GGC CAA TTC CGC TGC CGC GTG CCC GGC GCC TAC 477
 F F S F T V G K A P H K S L S V M L V R
 TTC TTC TCC TTC ACG GTT GGC AAG GCC CCG CAC AAG AGC CTG TCG GTG ATG CTG GTG CGG 537
 N H D E V Q A L A F D E Q R R P S A R R
 AAC CAC GAC GAG GTG CAG GCG CTG GCC TTC GAC GAG CAG AGG CGG CCC AGC GCA CGG CGC 597
 A A S Q S A M L Q L D Y G D T V W L R L
 GCC GCC AGC CAG AGC GCC ATG CTG CAG CTC GAC TAC GGC GAC ACA GTT TGG CTG CGG CTG 657
 H G A P Q Y A L G A P G A T F S G Y L V
 CAT GGC GCC CCG CAG TAC GCG CTG GGC GCG CCC GGC GCC ACC TTC AGC GGC TAC CTG GTC 717
 Y A D A D A D A P A R G P P A P P E P R
 TAC GCC GAC GCC GAC GCC GAC GCG CTT GCG CGC GGG CCG CCC GCG CCC CCC GAG CCG CGC 777
 S A F S A A R T R S L V G S D A G E G P
 TCG GCA TTC TCG GCG GCG CGC ACG CGC AGC CTG GTG GGC TCG GAC GCG GGG TCC GGG CCG 837
 R H R P L A F D T E L V N I G G D F D A
 CGG CAC CGG CCG CTA GCC TTC GAC ACC GAG CTC GTC AAC ATT GGC GGC GAC TTC GAC GCG 897
 A A G V F R C R L P G A Y F F S F T L G
 GCG GCC GGC GTG TTC CGC TGC CGC CTG CCC GGC GCC TAC TTC TTC TCC TTC ACG CTG GGC 957
 K L P R K T L S V K L M K N R D E V Q A
 AAG CTG CCG CGT AAG ACG CTG TCG GTG AAG CTG ATG AAG AAC CGC GAC GAG GTG CAG GCC 1017
 M I Y D D G A S R R R E M Q S Q S V M L
 ATG ATT TAC GAC GAC GGC GCG TCG CGG CGC CGC GAG ATG CAG AGC CAG AGC GTG ATG CTG 1077
 A L R R G D A V W L L S H D H D G Y G A
 GCC CTG CGG CGC GGC GAC GCC GTC TGG CTG CTC AGC CAC GAC CAC GAC GGC TAC GGC GCC 1137
 Y S N H G K Y I T F S G F L V Y P D L A
 TAC AGC AAC CAC GGC AAG TAC ATC ACC TTC TCC GGC TTC TTG GTG TAC CCC GAC CTC GCC 1197

FIGURE 25A

P A A P P G L G A P E E L
GCC GCC GCC CCG CCA GGC CTC GGG GCC CCG GAA CTC CTG TGA 1239

CCCCCGGCCAGAGAGGTGCCCCGGGAGGGCCAGGGGGCTGCATGCCGGGCCAGGCCCGGGCCCCGAACGCCTCGCGCGA 1318

CGCCACGGCCGCCGGGCGCGCCTGGACTCTGCCAATAAAGCAGAAAGCGGGCCTGCGCATTGCCCGGGC 1388

FIGURE 25B

	10	20	30	40	50	60	70
HUMAN	MLP	LL	GL	LG	PAAC	HAL	GPT
	10	20	30	40	50	60	70
MONKEY	MLP	LL	GL	LG	PAAC	HAL	GPT
	10	20	30	40	50	60	70
	80	90	100	110	120	130	140
	RV	P	G	A	Y	F	F
	80	90	100	110	120	130	140
	RV	P	G	A	Y	F	F
	150	160	170	180	190	200	210
	Q	Y	A	L	G	A	P
	150	160	170	180	190	200	210
	Q	Y	A	L	G	A	P
	220	230	240	250	260	270	280
	G	G	D	F	A	A	A
	220	230	240	250	260	270	280
	G	G	D	F	A	A	A
	290	300	310	320	330	340	
	G	D	A	V	W	L	S
	290	300	310	320	330	340	
	G	D	A	V	W	L	S

FIGURE 26

CONSENSUS *->AftvirstnrpPaEmsnpgqpViFdeVLyNqqghYdpaTGkFtCkvp
 AF++ r+c p++ + V Fd+V +N++g++d aTG F C vP
 31 AFSAARTT--PLE--GTSEMAVTFDKVYVNIGGDFDVATGQFRCRVP 73

CONSENSUS GLYyFsFhvsskgtRqnvcVsLmrSSrngvrqkVmeifcdeyakgtyqvas
 G Y Fsf + + +++ V L+r r+ v+ ++ f ++ g + +aS
 74 GAYFFSFTAGKAP-HKSLSVMLVR-NRDEVQ--ALAFDEQRRPGARRAAS 119

CONSENSUS GGavLqLrqGDrVWLelddkqctngllggegvhSvFSGFLl<-*
 +a LqL GD VWL l + + l ++ ++FSG+L+
 120 QSAMLQLDYGDTVWLRLHGAPQYALGAPG---ATFSGYLV 156

CONSENSUS *->AftvirstnrpPaEmsn.....pgqpViFdeVLyNqqghYdpaTGkF
 AF++ r+ + s+ +++++ qp Fd+ +N++g++d a G+F
 178 AFSAARTR-SLVG--SDagpgpRHQPLAFDTEFVNIGGDFDAAAGVF 221

CONSENSUS tCkvPGlYyFsFhvsskgtRqnvcVsLmrSSrngvrqkVmeifcdeyakgt
 C +pG Y Fsf + R+++ V Lm+ r+ v+ -m -d a
 222 RCRLPGAYFFSFTLGKLP-RKTLsvklmk-NRDEVQ--AMI-YDDGASRR 256

CONSENSUS yqvasGGavLqLrqGDrVWLelddkqctngl<-*
 ++ S ++ L+Lr+GD VWL + d +g+
 267 REMQSQSVMLALRRGDAVWLLSHDH--DGY 294

FIGURE 27

CONSENSUS ·->AftvirstnrpPaEmsnpgqpViFdeVLYNqqghYdpaTGkFtCkvP
 AF++ r+c p++ + V Fd+V +N++g++d aTG F C vP
 31 AFSAARTT--PLE--GASEMAVTFDKVYVNIGGDFDAATGQFRCRVP 73

CONSENSUS G1YyFsFhvsskgtRqnvcVsLmrSSrngvrqkVmefcdeyakgtyqvaS
 G Y FsF v + +++ V L+r + v+ ++ f ++ + +aS
 74 GAYFFSFTVVKAP-HKSLSVMLVR-NHDEVQ--ALAFDEQRRPSARRAAS 119

CONSENSUS GGavLqLrqGDrVWLeiddkqtnngllggegvhSvFSGFLl<-*
 +a LqL GD VWL i + + l ++ ++FSG+L+
 120 QSAMLQLDYGDTVWLRHLGAPQYALGAPG---ATFSGYLV 156

CONSENSUS ·->AftvirstnrpPaEmsn.....pgqpViFdeVLYNqqghYdpaTGkF
 AF++ r+ + s+ +++++ p Fd+ L+N++g++d a G+F
 178 AFSAARTR-SLVG--SDagsgpRHRPLAFDTELVNIGGDFDAAAGVF 221

CONSENSUS tCkvPG1YyFsFhvsskgtRqnvcVsLmrSSrngvrqkVmefcdeyakgt
 C +PG Y FsF + R+++ V Lm+ z+ v+ +m +d a
 222 RCRLPGAYFFSFTLGKLP-RKTLVKLMK-NRDEVQ--AMI-YDDGASRR 266

CONSENSUS yqvaSGGavLqLrqGDrVWLeiddkqtnngllggegvhSvFSGFLl<-*
 ++ S ++ L+Lr+GD VWL + d ++g ++ g++++FSGFL+
 267 REMQSQSVMLALRRGDAVWLLSHDHDGYGAYSNHGKYITFSGFLV 311

FIGURE 28

tcggaccgcc cgccaccagc cacgtgcc atg ctg ctg ctc ttg ctg ggc ttc	52
Met Leu Leu Leu Leu Leu Gly Phe	
1 5	
cta ggc ccg gcg gcc tgc tgg gca ctg ggc ccg gct ggc cct ggc tcc	100
Leu Gly Pro Ala Ala Cys Trp Ala Leu Gly Pro Ala Gly Pro Gly Ser	
10 15 20	
tcg gag ctg cgg tca gcc ttc tcg gcg gct cgc acc acc ccg ctg gag	148
Ser Glu Leu Arg Ser Ala Phe Ser Ala Ala Arg Thr Thr Pro Leu Glu	
25 30 35 40	
ggc acg tcg gag atg gcg gtg acc ttc gac aag gtg tac gtg aac atc	196
Gly Thr Ser Glu Met Ala Val Thr Phe Asp Lys Val Tyr Val Asn Ile	
45 50 55	
ggg ggt gac ttc gac gca gcc acc ggg cgg ttc cgc tgt cgc gtg ccg	244
Gly Gly Asp Phe Asp Ala Ala Thr Gly Arg Phe Arg Cys Arg Val Pro	
60 65 70	
ggc gcc tac ttc ttc tcc ttc acg gcc ggc aag gcc ccg cac aaa aac	292
Gly Ala Tyr Phe Phe Ser Phe Thr Ala Gly Lys Ala Pro His Lys Asn	
75 80 85	
ctg tcg gtg atg ctg gtg cgc aac cgc gac gaa gtg cag gcg ctg gct	340
Leu Ser Val Met Leu Val Arg Asn Arg Asp Glu Val Gln Ala Leu Ala	
90 95 100	
ttc gac aag cag cga cgg cca ggc gcg cgg cgc gcg gcc agc caa agc	388
Phe Asp Lys Gln Arg Arg Pro Gly Ala Arg Arg Ala Ala Ser Gln Ser	
105 110 115 120	
gcc atg ctg cag ctc gac tac ggc gac acg gtg tgg ctg cgg ctg cac	436
Ala Met Leu Gln Leu Asp Tyr Gly Asp Thr Val Trp Leu Arg Leu His	
125 130 135	
ggc gct ccg cat tac gcg ctc ggc gcg ccg ggc gcc acc ttc agc ggc	484
Gly Ala Pro His Tyr Ala Leu Gly Ala Pro Gly Ala Thr Phe Ser Gly	
140 145 150	
tac ctg gtg tac gcg gac gcc gac gcc gac gcg cct gcg cgc ggg ccc	532
Tyr Leu Val Tyr Ala Asp Ala Asp Ala Asp Ala Pro Ala Arg Gly Pro	
155 160 165	
gcg gcc ccg gag ccg cgc tcg gcc ttc tcc gcg cgc cac gcc acc tgg	580
Ala Ala Pro Glu Pro Arg Ser Ala Phe Ser Ala Arg His Ala Thr Trp	
170 175 180	
tgg gct ccg aac ccg ccc cgg ccc gcg cca cgg cgt ttg gcc ttc	625
Trp Ala Pro Asn Pro Pro Arg Pro Ala Pro Arg Arg Leu Ala Phe	
185 190 195	

FIGURE 29

GAP of: FrGcgManager_266_RYBH0yQG_ check: 1120 from: 1 to: 625

mM245 (analysis only) - Import - complete

to: FrGcgManager_266_SYB96sGK_ check: 9121 from: 1 to: 2747

hM245 (analysis only) - Import - complete

Gap Weight: 12 Average Match: 10.000
Length Weight: 4 Average Mismatch: 0.000

Quality: 5404 Length: 2749
Ratio: 8.646 Gaps: 10
Percent Similarity: 89.567 Percent Identity: 89.567

Match display thresholds for the alignment(s):

| = IDENTITY
: = 5
. = 1

FrGcgManager_266_RYBH0yQG_ x FrGcgManager_266_SYB96sGK_ May 30, 19100
15:01 ..

```

mouse 1 .....TCGGACCGCCCCGCCACCAGCCA 22
human 51 CAGATCGCGGGCTCAGACCAAACCCGACTC.GACCG.CCGCCCCCAGCCA 98
      23 CGTGCCATGCTGCTGCTCTTGCTGGGCTTCCTAGGCCCCGGCGGCCTGCTG 72
      99 GCGGCCATGCTGCCGCTTCTGCTGGGCCTGCTGGGCCCAGCGGCCTGCTG 148
      73 GGCACCTGGGCCCCG...GCTGGCCCTGGCTCCTCGGAGCTGCGGTCAGCCT 119
      149 GGCCCTGGGCCCCGACCCCGGGCCGGGATCCTCTGAGCTGCGCTCGGCCT 198
      120 TCTCGGCGGCTCGCACCAACCCGCTGGAGGGCACGTCGGAGATGGCGGTG 169
      199 TCTCGGCGGCACGCACCAACCCCTGGAGGGCACGTCGGAGATGGCGGTG 248
      170 ACCTTCGACAAGGTGTACGTGAACATCGGGGGTGAATTCGACGCAGCCAC 219
      249 ACCTTCGACAAGGTGTACGTGAACATCGGGGGCGACTTCGATGTGGCCAC 298
      220 CGGGCGGTTCCGCTGTCGCGTGCCGGGCGCCTACTTCTTCTCCTTCACGG 269
      299 CGGCCAGTTTCGCTGCCGCTGCCGGGCGCCTACTTCTTCTCCTTCACGG 348
      270 CCGGCAAGGCCCCGCACAAAACCTGTGCGGTGATGCTGGTGCGCAACCGC 319
      349 CTGGCAAGGCCCCGCACAAGAGCCTGTGCGGTGATGCTGGTGCGAAACCGC 398
      320 GACGAAGTGCAGGCGCTGGCTTTCGACAAGCAGCGACGGCCAGGCGCGCG 369
      399 GACGAGGTGCAGGCGCTGGCCTTCGACGAGCAGCGGCGGCCAGGCGCGCG 448

```

FIGURE 30A

mouse
 human

```

370 GCGCGCGGCCAGCCAAAGCGCCATGCTGCAGCTCGACTACGGCGACACGG 419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
449 GCGCGCAGCCAGCCAGAGCGCCATGCTGCAGCTCGACTACGGCGACACAG 498

420 TGTGGCTGCGGCTGCACGGCGCTCCGCATTACGCGCTCGGCGCGCCGGGC 469
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
499 TGTGGCTGCGGCTGCATGGCGCCCCGAGTACGCGCTAGGCGCGCCCGGC 548

470 GCCACCTTCAGCGGCTACCTGGTGTACGCGGACGCCGACGCCGACGCGCC 519
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
549 GCCACCTTCAGCGGCTACCTAGTCTACGCCGACGCCGACGCTGACGCGCC 598

520 TGC GCGCGG...GCCCCGCGGCCCGGAGCCGCGCTCGGCCTTCTCCGC.G 565
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
599 TGC GCGCGGGCCGCCCGCGCCCCCGAGCCGCGCTCGGCCTTCTCGGCGG 648

566 CGC.CAAGC.CA.CCTGGTGGGCTCCGA.ACCCGCCCCGGCCCGCGCCA. 610
    ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
649 CGCGCACGCGCAGCTTGGTGGGCTCGGACGCTGGCCCCGGGCCGCGGCAC 698

611 CGGCGTTTGGCCTTC..... 625
    | | | |||||
699 CAACCACTCGCCTTCGACACCGAGTTCGTCAACATTGGCGGCGACTTCGA 748
  
```

FIGURE 30B

```

      10      20      30      40      50
T246  MAA-QYGSMFNPSTPGASYGPG---RQEPRNSQLR----IVLVGKTGAGKSATGNSILGRKVFHS
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
AIG1  MANDQKNSESFPAKEDHKDDAAAPAEVDHKDEFSASQPHPVENIVLVGRTGNGKSATGNSIVRSKVFKS
      10      20      30      40      50      60      70

      60      70      80      90      100     110     120
GTAAKSITKKCEKRSSSWKETELV-VVDTPGIFDTEVPNAETSKETIIRCILLTSPGPHALLLVVPL-GRY
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
KTKSSGVTMECHAVKAVTPEGPILNVIDTPGLFDLSVSAEFIGKEIVKCLTLADGGLHAVLLVLSVRTRI
      80      90      100     110     120     130     140

      130     140     150     160     170     180     190
TEEEHKATEKILKMFGERRARSMILIFTRKDDLGDNT--LHDYLREAPEDIQDLMDIF-GDRYCALNNKA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
SQEEEMVLSTLQVLFGSKIVDYLIIVFTGGDVLEDDGMTLEDYLGDNMPDFLKRVLILCGQRMILFDNKT
      150     160     170     180     190     200     210

      200     210     220     230     240     250     260
TGAEQEAQRAQ-LLGLIQRVVRENKEGCTNRMYQRAEEEEIQKQTQAMQELH-RVELEREKARIREEYEE
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
KDDEKKTQVHELLKLIDLVRKQNNNIPYTDMEYHMIKEENERHHKKEQEELESKGHSEEQLAALMKELQI
      220     230     240     250     260     270     280

      270     280     290     300     310     320
KI-RKLEDKVEQ-EKRKK-QME--KKLAEQEAHYAVRQQRARTEVESKDGIELIMTALQIASFILLRLF
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
MNERNLKAMAEMMEKNMKIAMEAQEKLFEQREKAQEMSYYQKMEMQEKLKQMEGRMRAEMEAQMLSRQQC
      290     300     310     320     330     340     350

AED
.
SIL

```

FIGURE 31

```

      40      50      60
inputs MRQAGGLGLLALLLLALLLGPGRGVGRPGSGAQAGRWAQRFKVVFAPVICKRTCLKGQ
      .....
      MRQAA-LGLLALLLLALLLGPGRGVGRPGSGAQAGRWAQRFKVVFAPVICKRTCLKGQ
      10      20      30      40      50

      70      80      90      100      110      120
inputs CRDSCQQGSNMTLIGENGHSTDTLTGSAFRVVVCP L P C M N G G Q C S S R N Q C L C P P D F T G R F
      .....
      CRDSCQQGSNMTLIGENGHSTDTLTGSAFRVVVCP L P C M N G G Q C S S R N Q C L C P P D F T G R F
      60      70      80      90      100      110

      130      140      150      160      170      180
inputs CQVPAAGTGAGTGSSSGPWPDRAMSTG L P L P L A P E G E S V A S K H A I Y A V Q V I A D P P G P G E G
      .....
      CQVPAAGTGAGTGSSSGPWPDRAMSTG L P L P L A P E G E S V A S K H A I Y A V Q V I A D P P G P G E G
      120      130      140      150      160      170

      190      200      210      220      230      240
inputs PPAQHA AFLVPLGPGQISAEVQAPPVVNVRVHHPPEASVQVHRIEGPNAEGPASSQHLL
      .....
      PPAQHA AFLVPLGPGQISAEVQAPPVVNVRVHHPPEASVQVHRIEGPNAEGPASSQHLL
      180      190      200      210      220      230

      250      260      270      280      290      300
inputs PHPKPPHPRPPTQKPLGRCFQDTLPKQPCGSNPLPGLTKQEDCCSIGTAWGQSKCHKCP
      .....
      PHPKPPHPRPPTQKPLGRCFQDTLPKQPCGSNPLPGLTKQEDCCSIGTAWGQSKCHKCP
      240      250      260      270      280      290

      310      320      330      340      350      360
inputs QLQYTG VQKPVVVRGEVGADCPQGYKRLNSTHCQDINECAMPGNVCHGDC L N N P G S Y R C V
      .....
      QLQYTG VQKPVVVRGEVGADCPQGYKRLNSTHCQDINECAMPGNVCHGDC L N N P G S Y R C V
      300      310      320      330      340      350

      370      380      390      400      410      420
inputs CPPGHS LGPLAAQCIADKPEEKSLCFRLVSTEHQCQHPLTTRLTRQLCCCSVGKAWGARC
      .....
      CPPGHS LGPLAAQCIADKPEEKSLCFRLVSTEHQCQHPLTTRLTRQLCCCSVGKAWGARC
      360      370      380      390      400      410

      430      440      450      460      470      480
inputs QRCPADGTAAFKEICPGWERVPYPHLP P D A H H P G G K R L L P L P A P D G P P K P Q O L P E S P S R A
      .....
      QRCPADGTAAFKEICPGWERVPYPHLP P D A H H P G G K R L L P L P A P D G P P K P Q O L P E S P S R A
      420      430      440      450      460      470

      490      500      510      520      530      540
inputs PPLEDTEEEERGVTMDPPVSEERSVQQSHPTTTTSPRPYPPELISRPSPTTFHRFLPDLPP
      .....
      PPLEDTEEEERGVTMDPPVSEERSVQQSHPTTTTSPRPYPPELISRPSPTTFHRFLPDLPP
      480      490      500      510      520      530

      550      560      570      580      590      600
inputs SRSAVEIAPTQVTETDECRLNQNICGHGQCVPGPSDYSCHCNAGYRSHPOHRYCVDVNEC
      .....
      SRSAVEIAPTQVTETDECRLNQNICGHGQCVPGPSDYSCHCNAGYRSHPOHRYCVDVNEC
      540      550      560      570      580      590

      610      620      630      640      650      660
inputs EAEP CGPGKGICMNTGGSYNCHCNRGYRLHVGAGGRSCVDLNECAKPHLCGDDGGFCINFP
      .....
      EAEP CGPGKGICMNTGGSYNCHCNRGYRLHVGAGGRSCVDLNECAKPHLCGDDGGFCINFP
      600      610      620      630      640      650

```

59/60

```

        670      680      690      700      710      720
inputs GHYKNCYPGYRLKASRPPICEDIDECRDPSTCPDGKCNKPGSFKCIACQPGYRSQGGG
        660      670      680      690      700      710
        GHYKNCYPGYRLKASRPPICEDIDECRDPSTCPDGKCNKPGSFKCIACQPGYRSQGGG

        730      740      750      760      770      780
inputs ACRDVNECSEGTPCSPGWCENLPGSYRCTCAQGI RTRTGR LSCIDVDECEAGKVCQDGIC
        720      730      740      750      760      770
        ACRDVNECSEGTPCSPGWCEKLP GSYRCTCAQGI RTRTGR LSCIDVDDCEAGKVCQDGIC

        790      800      810      820      830      840
inputs TNTPGSFQCQLSGYHLSRDRSRCEDIDECDFPAACIGGDCINTNGSYRCLCPLGHR LVG
        780      790      800      810      820      830
        TNTPGSFQCQLSGYHLSRDRSRCEDIDECDFPAACIGGDCINTNGSYRCLCPLGHR LVG

        850      860      870      880      890      900
inputs GRKCKKDIDEC SQDPGLCLPHACENLQGSYVCV CDEGFTLTQDQHGC EEVEQPHHKKECY
        840      850      860      870      880      890
        GRKCKKDIDEC SQDPGLCLPHACENLQGSYVCV CDEGFTLTQDQHGC EEVEQPHHKKECY

        910      920      930      940      950      960
inputs LNFDDTVFCDSVLATNV TQQECCCSLGAGWGDHCEIYPCPVYSSAEFHS LVPD GKRLHSG
        900      910      920      930      940      950
        LNFDDTVFCDSVLATNV TQQECCCSLGAGWGDHCEIYPCPVYSSAEFHS LVPD GKRLHSG

        970      980      990      1000      1010      1020
inputs QQHCELCIPAHRDIDEC ILFGAEICKEGKCVNTQPGYECYCKQGFYYDGNLLECVDVDEC
        960      970      980      990      1000      1010
        QQHCELCIPAHRDIDEC ILFGAEICKEGKCVNSQPGYECYCKQGFYYDGNLLECVDVDEC

        1030      1040      1050      1060      1070
inputs LDESNCRN GVCENTRGGYRC ACTPPAEYSPAQAQCLIPERWS-TPQRDVKCAGASEERTA
        1020      1030      1040      1050      1060      1070
        LDESNCRN GVCENTWR-LPCA CTPPAEYSPAQAQCL SPEEMEHAPERREVCWGQRGEDGM

        1080      1090      1100      1110      1120      1130
inputs CVWGPWAGPALTFDDCCCRQ PRLGTQCRPCPPRG TGSQCPTSQSESNSFWDT SPLLLGKS
        1080      1090      1100      1110      1120      1130
        CM-GPLAGPALTFDDCCCRQ PRLGYQCRPCPPRG TGSQCPTSQSESNSFWDT SPLLLGKS

        1140      1150      1160      1170      1180      1190
inputs PRDEDSSEEDSDECR CVSGRCVPRPGGAVCECPGGFQLDASRARCVDI DECRELNQRGLL
        1140      1150      1160      1170      1180      1190
        PRDEDSSEEDSDECR CVSGPCVPRPGGAVCECPGGFQLDASRARCVDI DECRELNQRGLL

        1200      1210      1220      1230      1240      1250
inputs CKSERC VNTSGSFRCVCKAGFTRSRPHGPACLSAAADDAI AHTSVIDHRGYFH
        1200      1210      1220      1230      1240      1250
        CKSERC VNTSGSFRCVCKAGFTRSRPHGPACLSAAADDAI AHTSVIDHRGYFH

```

FIGURE 32.B

SEQUENCE LISTING

<110> Millennium Pharmaceuticals, Inc.

<120> SECRETED PROTEINS AND USES THEREOF

<130> 7853-205-228

<140>

<141>

<150> 09/342,687

<151> 1999-06-29

<160> 93

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 1513

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (85)...(570)

<221> modified_base

<222> all "n" positions

<223> n = a, c, g, or t

<400> 1

```

gtcgacccac gcgctccggcc cgacacccag accgaccttg accgcccacc tggcaggagc      60
aggacaggac ggccggacgc ggcc atg gcc gag ctc ccg ggg ccc ttt ctc      111
                Met Ala Glu Leu Pro Gly Pro Phe Leu
                   1                   5

tgc ggg gcc ctg cta ggc ttc ctg tgc ctg agt gtt ccc ccc agt aat      159
Cys Gly Ala Leu Leu Gly Phe Leu Cys Leu Ser Val Pro Pro Ser Asn
  10                   15                   20                   25

ccc tta tgc agt cag agt gga caa acc tct gtg gga ggc tct act gca      207
Pro Leu Cys Ser Gln Ser Gly Gln Thr Ser Val Gly Gly Ser Thr Ala
                   30                   35                   40

ctg aga tgc agc tct tcc gag ggg gct cct aag cca gtg tac aac tgg      255
Leu Arg Cys Ser Ser Ser Glu Gly Ala Pro Lys Pro Val Tyr Asn Trp
                   45                   50                   55

gtg cgt ctt gga act ttt cct aca cct tct cct ggc agc atg gtt caa      303
Val Arg Leu Gly Thr Phe Pro Thr Pro Ser Pro Gly Ser Met Val Gln
                   60                   65                   70

gat gag gtg tct ggc cag ctc att ctc acc aac ctc tcc ctg acc tcc      351
Asp Glu Val Ser Gly Gln Leu Ile Leu Thr Asn Leu Ser Leu Thr Ser
                   75                   80                   85

```

```

tcg ggc acc tac cgc tgt gtg gcc acc aac cag atg ggc agt gca tcc      399
Ser Gly Thr Tyr Arg Cys Val Ala Thr Asn Gln Met Gly Ser Ala Ser
  90                      95                      100                      105

tgt gag ctg acc ctc tct gtg acc gaa ccc tcc caa ggc cga gtg gcc      447
Cys Glu Leu Thr Leu Ser Val Thr Glu Pro Ser Gln Gly Arg Val Ala
                110                      115                      120

gga gct ctg att ggg gtg ctc ctg ggc gtg ctg ttg ctg tca gtt gct      495
Gly Ala Leu Ile Gly Val Leu Leu Gly Val Leu Leu Leu Ser Val Ala
                125                      130                      135

gcg ttc tgc ctg gtc agg ttc cag aaa gag agg ggg aag aag ccc aag      543
Ala Phe Cys Leu Val Arg Phe Gln Lys Glu Arg Gly Lys Lys Pro Lys
    140                      145                      150

gag aca tat ggg ggt agt gac ctt cgg tgagcaggag ggctgggggg          590
Glu Thr Tyr Gly Gly Ser Asp Leu Arg
    155                      160

tggcgcaagg agggaggaaa gggcttgagt taaaagcggg tgccctgcaac cctcaaactc      650
cgacatcatt cagtgtgttt aggggcagga ggtgttggtc agccgtggaa tttgctggtg      710
gcagcagtgt aacctgtgta tttgagggtg caggcaagcg gtacagggtg gagtggctgg      770
tccacaagct gtggcaggga agctgtttgc aggactgcc tgccccctcct catatttaat      830
aaagtttact tttctgttcc gaaggatatt tcatatattt taaccacctg ggagtagtag      890
tggcttgtag atgccaggaa atggatttgt cctgagcagt cagctgagtt caattcttct      950
gtggaggaaa tcaggaaagg ggaggggaaa ctgcctctgt catccacttt agctgccagn     1010
caggggtctag gatagggatc agagcaacat ttcttcagggt ggagtcctca gattacctgg     1070
acagaaatca ccgggaacta gttatacatt cagattcagg ccacttctag ccttctctgta     1130
gttgtgcgtt ggggagtgat nagggccana aatttcnttt taaccaaagt tccncanatt     1190
atittcaagc ccagtgaat ttaagagtcc ccaggttaga ggacggccct ccnccgcagg     1250
aggnttttac tgkttactca gaacttgct ataccatca gggaggatgc catcgctcct      1310
gggatctctg agcacacttg tatgagggt gattctagca aggggttctt ggaaagaccc      1370
tcgtctgcca gcaccgtgac gaccaccaag tccaagctcc ctatggtcgt gtgacttctc      1430
ccgatccctg agggcgggta gggggaatat caataattaa agtctgtggg taccaaaaaa      1490
aaaaaaaaaa aaagggcggc cgc                                              1513

```

<210> 2
 <211> 162
 <212> PRT
 <213> Homo sapiens

```

<400> 2
Met Ala Glu Leu Pro Gly Pro Phe Leu Cys Gly Ala Leu Leu Gly Phe
  1          5          10          15
Leu Cys Leu Ser Val Pro Pro Ser Asn Pro Leu Cys Ser Gln Ser Gly
    20          25          30
Gln Thr Ser Val Gly Gly Ser Thr Ala Leu Arg Cys Ser Ser Ser Glu
    35          40          45
Gly Ala Pro Lys Pro Val Tyr Asn Trp Val Arg Leu Gly Thr Phe Pro
    50          55          60
Thr Pro Ser Pro Gly Ser Met Val Gln Asp Glu Val Ser Gly Gln Leu
    65          70          75          80
Ile Leu Thr Asn Leu Ser Leu Thr Ser Ser Gly Thr Tyr Arg Cys Val
    85          90          95
Ala Thr Asn Gln Met Gly Ser Ala Ser Cys Glu Leu Thr Leu Ser Val
    100          105          110

```

Thr Glu Pro Ser Gln Gly Arg Val Ala Gly Ala Leu Ile Gly Val Leu
 115 120 125
 Leu Gly Val Leu Leu Leu Ser Val Ala Ala Phe Cys Leu Val Arg Phe
 130 135 140
 Gln Lys Glu Arg Gly Lys Lys Pro Lys Glu Thr Tyr Gly Gly Ser Asp
 145 150 155 160
 Leu Arg

<210> 3
 <211> 486
 <212> DNA
 <213> Homo sapiens

<400> 3
 atggccgagc tcccgggggcc ctttctctgc ggggccctgc taggcttccct gtgcctgagt 60
 gttcccccca gtaatccctt atgcagtcag agtggacaaa cctctgtggg aggcctact 120
 gcactgagat gcagctcttc cgaggggggt cctaagccag tgtacaactg ggtgcgtctt 180
 ggaacttttc ctacaccttc tcctggcagc atggttcaag atgaggtgtc tggccagctc 240
 attctcacca acctctccct gacctcctcg ggcacctacc gctgtgtggc caccaaccag 300
 atgggcagtg catcctgtga gctgacctc tctgtgaccg aacctcccca aggccgagtg 360
 gccggagctc tgattggggg gctcctgggc gtgctgttgc tgtcagttgc tgcgttctgc 420
 ctggtcaggt tccagaaaga gagggggaag aagcccaagg agacatatgg gggtagtgac 480
 cttcgg 486

<210> 4
 <211> 1992
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (94) ... (1080)

<400> 4
 gtcgacccac gcgtccgcaa catcctggct tagtattgtg tgcaaaatca gagaggggtg 60
 caagatcctg atttttcagg agttcaagcg aca atg gca gcc caa tac ggc agt 114
 Met Ala Ala Gln Tyr Gly Ser
 1 5
 atg agc ttc aac ccc agc aca cca ggg gcc agt tat ggg cct gga agg 162
 Met Ser Phe Asn Pro Ser Thr Pro Gly Ala Ser Tyr Gly Pro Gly Arg
 10 15 20
 caa gag ccc aga aat tcc caa ttg aga att gtg tta gtg ggt aaa acc 210
 Gln Glu Pro Arg Asn Ser Gln Leu Arg Ile Val Leu Val Gly Lys Thr
 25 30 35
 gga gca gga aaa agt gca aca gga aac agc atc ctt ggc cgg aaa gtg 258
 Gly Ala Gly Lys Ser Ala Thr Gly Asn Ser Ile Leu Gly Arg Lys Val
 40 45 50 55
 ttt cat tct ggc act gca gca aaa tcc att acc aag aag tgt gag aaa 306
 Phe His Ser Gly Thr Ala Ala Lys Ser Ile Thr Lys Lys Cys Glu Lys
 60 65 70

cgc agc agc tca tgg aag gaa aca gaa ctt gtc gta gtt gac aca cca	354
Arg Ser Ser Ser Trp Lys Glu Thr Glu Leu Val Val Val Asp Thr Pro	
75 80 85	
ggc att ttc gac aca gag gtg ccc aat gct gaa acg tcc aag gag att	402
Gly Ile Phe Asp Thr Glu Val Pro Asn Ala Glu Thr Ser Lys Glu Ile	
90 95 100	
att cgc tgc att ctt ctg acc tcc cca ggg cct cat gct ctg ctt ctg	450
Ile Arg Cys Ile Leu Leu Thr Ser Pro Gly Pro His Ala Leu Leu Leu	
105 110 115	
gtg gtt cca ctg ggc cgt tac act gag gaa gag cac aaa gcc aca gag	498
Val Val Pro Leu Gly Arg Tyr Thr Glu Glu Glu His Lys Ala Thr Glu	
120 125 130 135	
aag atc ctg aaa atg ttt gga gag agg gct aga agt ttc atg att ctc	546
Lys Ile Leu Lys Met Phe Gly Glu Arg Ala Arg Ser Phe Met Ile Leu	
140 145 150	
ata ttc acc cgg aaa gat gac tta ggt gac acc aat ttg cat gac tac	594
Ile Phe Thr Arg Lys Asp Asp Leu Gly Asp Thr Asn Leu His Asp Tyr	
155 160 165	
tta agg gaa gct cca gaa gac att caa gac ttg atg gac att ttc ggt	642
Leu Arg Glu Ala Pro Glu Asp Ile Gln Asp Leu Met Asp Ile Phe Gly	
170 175 180	
gac cgc tac tgt gcg tta aac aac aag gca aca ggc gct gag cag gag	690
Asp Arg Tyr Cys Ala Leu Asn Asn Lys Ala Thr Gly Ala Glu Gln Glu	
185 190 195	
gcc cag agg gca cag ttg ctg ggc ctg atc cag cgc gtg gtg agg gag	738
Ala Gln Arg Ala Gln Leu Leu Gly Leu Ile Gln Arg Val Val Arg Glu	
200 205 210 215	
aac aag gaa ggc tgc tac act aat agg atg tac caa agg gcg gag gag	786
Asn Lys Glu Gly Cys Tyr Thr Asn Arg Met Tyr Gln Arg Ala Glu Glu	
220 225 230	
gag atc cag aag caa aca caa gca atg caa gaa ctc cac aga gtg gag	834
Glu Ile Gln Lys Gln Thr Gln Ala Met Gln Glu Leu His Arg Val Glu	
235 240 245	
ctg gag aga gag aaa gcg cgg ata aga gag gag tat gaa gag aaa atc	882
Leu Glu Arg Glu Lys Ala Arg Ile Arg Glu Glu Tyr Glu Glu Lys Ile	
250 255 260	
aga aag ctg gaa gat aaa gtg gag cag gaa aag aga aag aag caa atg	930
Arg Lys Leu Glu Asp Lys Val Glu Gln Glu Lys Arg Lys Lys Gln Met	
265 270 275	
gag aag aaa cta gca gaa cag gag gct cac tat gct gta agg cag caa	978
Glu Lys Lys Leu Ala Glu Gln Glu Ala His Tyr Ala Val Arg Gln Gln	
280 285 290 295	

agg gca aga acg gaa gtg gag agt aag gat ggg ata ctt gaa tta atc 1026
 Arg Ala Arg Thr Glu Val Glu Ser Lys Asp Gly Ile Leu Glu Leu Ile
 300 305 310

atg aca gcg tta cag att gct tcc ttt att ttg tta cgt ctg ttc gcg 1074
 Met Thr Ala Leu Gln Ile Ala Ser Phe Ile Leu Leu Arg Leu Phe Ala
 315 320 325

gaa gat taaacttaat gaaaatctgt ttgtattttc tgcattattct ctggcaacct 1130
 Glu Asp

tgccccatac ttacttattt agcatagtcg agtgctctag tttctgtctc tcaggcactc 1190
 gtaactaagg accaccattg gccattggta gatgtttgat tgacttaaca agagagggac 1250
 aaattttcaa tttgtgaaac tccaaagcag aaagtattgg tgcttgctac cttgtgaatt 1310
 cttccttaga catgcagaga aaatgtatgc aagagaccaa aaagatggct ccaagctatg 1370
 tcatgttacc tgtaataaaa tcttttcttc tagattcttt ctatgttggc agataatctc 1430
 cccttgtagc ttccactcac ttattcttgc attcagagtc acaatgatca tcttaccat 1490
 gtgggtttttg agaaagaaaag atcaattctt tgtttgcagt aggtaatctt agagatggag 1550
 atgattgtag aattattcct agatgagtgat caatttattt aattccattg tcatataagg 1610
 agtcaaattg tttcttatca tttgttcatt gaagaacaga gacctgtctg gaaaatcgat 1670
 ctctacaaat tcaattaaat aatgatcccc aaatgctgaa aaagtgaaat acagcaattc 1730
 aacagataat agagcaatgt ttagtatatt cagctgtatc tgtagaaact ctttgacgaa 1790
 cctcaattta accaatttga tgaataccca gttctcttct tttctagaga aagatagttg 1850
 caacctcacc tccctcactc aacactttga atacttattg tttggcagggt catccacaca 1910
 cttctgcccc cactgcattg aattttttgc ttatgttggt tataataaaa cttttcaatt 1970
 atctcaaaaa aaaaaaaaaa aa 1992

<210> 5

<211> 329

<212> PRT

<213> Homo sapiens

<400> 5

Met Ala Ala Gln Tyr Gly Ser Met Ser Phe Asn Pro Ser Thr Pro Gly
 1 5 10 15
 Ala Ser Tyr Gly Pro Gly Arg Gln Glu Pro Arg Asn Ser Gln Leu Arg
 20 25 30
 Ile Val Leu Val Gly Lys Thr Gly Ala Gly Lys Ser Ala Thr Gly Asn
 35 40 45
 Ser Ile Leu Gly Arg Lys Val Phe His Ser Gly Thr Ala Ala Lys Ser
 50 55 60
 Ile Thr Lys Lys Cys Glu Lys Arg Ser Ser Ser Trp Lys Glu Thr Glu
 65 70 75 80
 Leu Val Val Val Asp Thr Pro Gly Ile Phe Asp Thr Glu Val Pro Asn
 85 90 95
 Ala Glu Thr Ser Lys Glu Ile Ile Arg Cys Ile Leu Leu Thr Ser Pro
 100 105 110
 Gly Pro His Ala Leu Leu Leu Val Pro Leu Gly Arg Tyr Thr Glu
 115 120 125
 Glu Glu His Lys Ala Thr Glu Lys Ile Leu Lys Met Phe Gly Glu Arg
 130 135 140
 Ala Arg Ser Phe Met Ile Leu Ile Phe Thr Arg Lys Asp Asp Leu Gly
 145 150 155 160
 Asp Thr Asn Leu His Asp Tyr Leu Arg Glu Ala Pro Glu Asp Ile Gln
 165 170 175
 Asp Leu Met Asp Ile Phe Gly Asp Arg Tyr Cys Ala Leu Asn Asn Lys
 180 185 190

Ala Thr Gly Ala Glu Gln Glu Ala Gln Arg Ala Gln Leu Leu Gly Leu
 195 200 205
 Ile Gln Arg Val Val Arg Glu Asn Lys Glu Gly Cys Tyr Thr Asn Arg
 210 215 220
 Met Tyr Gln Arg Ala Glu Glu Glu Ile Gln Lys Gln Thr Gln Ala Met
 225 230 235 240
 Gln Glu Leu His Arg Val Glu Leu Glu Arg Glu Lys Ala Arg Ile Arg
 245 250 255
 Glu Glu Tyr Glu Glu Lys Ile Arg Lys Leu Glu Asp Lys Val Glu Gln
 260 265 270
 Glu Lys Arg Lys Lys Gln Met Glu Lys Lys Leu Ala Glu Gln Glu Ala
 275 280 285
 His Tyr Ala Val Arg Gln Gln Arg Ala Arg Thr Glu Val Glu Ser Lys
 290 295 300
 Asp Gly Ile Leu Glu Leu Ile Met Thr Ala Leu Gln Ile Ala Ser Phe
 305 310 315 320
 Ile Leu Leu Arg Leu Phe Ala Glu Asp
 325

<210> 6
 <211> 987
 <212> DNA
 <213> Homo sapiens

<400> 6
 atggcagccc aatacggcag tatgagcttc aaccccagca caccaggggc cagttatggg 60
 cctggaaggc aagagcccag aaattcccaa ttgagaattg tgtagtgagg taaaaccgga 120
 gcaggaaaaa gtgcaacagg aaacagcatc cttggccgga aagtgtttca ttctggcact 180
 gcagcaaaat ccattaccaa gaagtgtgag aaacgcagca gtcctatggaa ggaaacagaa 240
 cttgtcgtag ttgacacacc aggcattttc gacacagagg tgcccaatgc tgaaacgtcc 300
 aaggagatta ttcgctgcat tttcttgacc tccccagggc ctcatgctct gcttctggtg 360
 gttccactgg gccgttacac tgaggaagag caciaagcca cagagaagat cctgaaaatg 420
 tttggagaga gggctagaag tttcatgatt ctcatattca cccggaaaga tgacttaggt 480
 gacaccaatt tgcctgacta ctttaaggga gctccagaag acattcaaga cttgatggac 540
 attttcgggtg accgctactg tgcgttaaac aacaaggcaa caggcgtgga gcaggaggcc 600
 cagagggcac agttgctggg cctgatccag cgcgtggtga gggagaacaa ggaaggctgc 660
 tacactaata ggatgtacca aagggcggag gaggagatcc agaagcaaac acaagcaatg 720
 caagaactcc acagagtggg gctggagaga gagaaagcgc ggataagaga ggagtatgaa 780
 gagaaaatca gaaagctgga agataaagtg gagcaggaaa agagaaagaa gcaaatggag 840
 aagaaactag cagaacagga ggctcactat gctgtaaggc agcaaagggc aagaacggaa 900
 gtggagagta aggatgggat acttgaatta atcatgacag cgttacagat tgcttccttt 960
 attttgttac gtctgttcgc ggaagat 987

<210> 7
 <211> 4225
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (65)...(3931)

<400> 7
 gtcgacccac gcgtccgccc ggatgcccg gccccaggg gctgctggcg gcctggcccc 60
 tgag atg cgc ggg gcg ggg gcg gcg ggg ctg ctg gcg ctg ctg ctg ctg 109
 Met Arg Gly Ala Gly Ala Ala Gly Leu Leu Ala Leu Leu Leu Leu
 1 5 10 15

ctg ctg ctg ctg ctg ctg ggc ctg ggc ggc agg gtc gag ggg ggg ccg	157
Leu Leu Leu Leu Leu Leu Gly Leu Gly Gly Arg Val Glu Gly Gly Pro	
20 25 30	
gcc ggc gag cgg ggc gca ggc ggg ggc ggg gcg ctg gcc cgc gag cgc	205
Ala Gly Glu Arg Gly Ala Gly Gly Gly Ala Leu Ala Arg Glu Arg	
35 40 45	
ttc aag gtg gtc ttt gcg ccg gtg atc tgc aag cgg acc tgt ctc aag	253
Phe Lys Val Val Phe Ala Pro Val Ile Cys Lys Arg Thr Cys Leu Lys	
50 55 60	
ggc cag tgt cgg gac agt tgt cag cag ggc tcc aac atg acg ctc atc	301
Gly Gln Cys Arg Asp Ser Cys Gln Gln Gly Ser Asn Met Thr Leu Ile	
65 70 75	
gga gag aac ggc cac agc aca gac acg ctc acg ggc tcc ggc ttc cgc	349
Gly Glu Asn Gly His Ser Thr Asp Thr Leu Thr Gly Ser Gly Phe Arg	
80 85 90 95	
gtg gtg gtg tgc cct ctc ccc tgc atg aat ggc ggc cag tgc tcc tcg	397
Val Val Val Cys Pro Leu Pro Cys Met Asn Gly Gly Gln Cys Ser Ser	
100 105 110	
cga aac cag tgc ctg tgt ccc ccg gac ttc act ggg cgc ttc tgc cag	445
Arg Asn Gln Cys Leu Cys Pro Pro Asp Phe Thr Gly Arg Phe Cys Gln	
115 120 125	
gtg ccc gca gga gga gcc ggt ggg ggt acc ggc ggc tca ggc ccc ggc	493
Val Pro Ala Gly Gly Ala Gly Gly Gly Thr Gly Gly Ser Gly Pro Gly	
130 135 140	
ctg agc agg aca ggg gcc ctg tcc aca ggg gcg ctg ccg ccc ctg gct	541
Leu Ser Arg Thr Gly Ala Leu Ser Thr Gly Ala Leu Pro Pro Leu Ala	
145 150 155	
ccg gag ggc gac tct gtg gcc agc aag cac gcc atc tac gcc gtc cag	589
Pro Glu Gly Asp Ser Val Ala Ser Lys His Ala Ile Tyr Ala Val Gln	
160 165 170 175	
gtg atc gct gac cct cct ggg ccc ggg gag ggg cct cct gcc cag cac	637
Val Ile Ala Asp Pro Pro Gly Pro Gly Glu Gly Pro Pro Ala Gln His	
180 185 190	
gca gcc ttc ctg gtg ccc cta ggc ccg gga cag atc tca gca gaa gtg	685
Ala Ala Phe Leu Val Pro Leu Gly Pro Gly Gln Ile Ser Ala Glu Val	
195 200 205	
cag gcc ccg ccc ccc gtg gtg aat gtg cgc gtc cat cac ccg ccc gag	733
Gln Ala Pro Pro Pro Val Val Asn Val Arg Val His His Pro Pro Glu	
210 215 220	
gcc tca gtc cag gtg cac cgc att gag agc tcg aac gcc gag agc gca	781
Ala Ser Val Gln Val His Arg Ile Glu Ser Ser Asn Ala Glu Ser Ala	
225 230 235	

gcc ccc tcc cag cac ctg ctg ccg cac ccc aag ccc tcg cac cct cgg	829
Ala Pro Ser Gln His Leu Leu Pro His Pro Lys Pro Ser His Pro Arg	
240 245 250 255	
ccg ccc acc cag aag ccc ctg ggc cgc tgc ttt cag gac act ctg ccc	877
Pro Pro Thr Gln Lys Pro Leu Gly Arg Cys Phe Gln Asp Thr Leu Pro	
260 265 270	
aag cag ccg tgt ggc agc aac ccc ctc ccc ggc ctc acc aag cag gaa	925
Lys Gln Pro Cys Gly Ser Asn Pro Leu Pro Gly Leu Thr Lys Gln Glu	
275 280 285	
gac tgc tgc ggt agc atc ggc act gcc tgg ggc cag agc aag tgc cac	973
Asp Cys Cys Gly Ser Ile Gly Thr Ala Trp Gly Gln Ser Lys Cys His	
290 295 300	
aag tgt ccc cag ctg cag tac aca gga gtg cag aag cca ggg cct gta	1021
Lys Cys Pro Gln Leu Gln Tyr Thr Gly Val Gln Lys Pro Gly Pro Val	
305 310 315	
cgt ggg gaa gtg ggc gct gac tgt ccc cag ggc tac aag agg ctt aac	1069
Arg Gly Glu Val Gly Ala Asp Cys Pro Gln Gly Tyr Lys Arg Leu Asn	
320 325 330 335	
agc acc cac tgc cag gac atc aac gag tgc gca atg ccg ggc gtg tgt	1117
Ser Thr His Cys Gln Asp Ile Asn Glu Cys Ala Met Pro Gly Val Cys	
340 345 350	
cgc cat ggt gac tgc ctc aac aac cct ggc tcc tat cgc tgt gtc tgc	1165
Arg His Gly Asp Cys Leu Asn Asn Pro Gly Ser Tyr Arg Cys Val Cys	
355 360 365	
cca cct ggc cat agt tta ggc ccc tcc cgt aca cag tgc att gca gac	1213
Pro Pro Gly His Ser Leu Gly Pro Ser Arg Thr Gln Cys Ile Ala Asp	
370 375 380	
aaa ccg gag gag aag agc ctg tgt ttc cgc ctg gtg agc cct gag cac	1261
Lys Pro Glu Glu Lys Ser Leu Cys Phe Arg Leu Val Ser Pro Glu His	
385 390 395	
cag tgc cag cac cca ctg acc acc cgc ctg acc cgc cag ctc tgc tgc	1309
Gln Cys Gln His Pro Leu Thr Thr Arg Leu Thr Arg Gln Leu Cys Cys	
400 405 410 415	
tgc agt gtc ggc aag gcc tgg ggc gcg cgg tgt cag cgc tgc cca aca	1357
Cys Ser Val Gly Lys Ala Trp Gly Ala Arg Cys Gln Arg Cys Pro Thr	
420 425 430	
gat ggc acc gct gcg ttc aag gag atc tgc cca gct ggg aag gga tac	1405
Asp Gly Thr Ala Ala Phe Lys Glu Ile Cys Pro Ala Gly Lys Gly Tyr	
435 440 445	
cac att ctc acc tcc cac cag acg ctc acc att cag ggc gag agt gac	1453
His Ile Leu Thr Ser His Gln Thr Leu Thr Ile Gln Gly Glu Ser Asp	
450 455 460	

ttt tcc ctt ttc ctg cac cct gac ggg cca ccc aag ccc cag cag ctt	1501
Phe Ser Leu Phe Leu His Pro Asp Gly Pro Pro Lys Pro Gln Gln Leu	
465 470 475	
ccg gag agc cct agc cag gct cca cca cct gag gac aca gag gaa gag	1549
Pro Glu Ser Pro Ser Gln Ala Pro Pro Pro Glu Asp Thr Glu Glu Glu	
480 485 490 495	
aga ggg gtg acc acg gac tca ccg gtg agt gag gag agg tca gtg cag	1597
Arg Gly Val Thr Thr Asp Ser Pro Val Ser Glu Glu Arg Ser Val Gln	
500 505 510	
cag agc cac cca act gcc acc acg act cct gcc cgg ccc tac ccc gag	1645
Gln Ser His Pro Thr Ala Thr Thr Pro Ala Arg Pro Tyr Pro Glu	
515 520 525	
ctg atc tcc cgt ccc tcg ccc ccg acc atg cgc tgg ttc ctg ccg gac	1693
Leu Ile Ser Arg Pro Ser Pro Pro Thr Met Arg Trp Phe Leu Pro Asp	
530 535 540	
ttg cct cct tcc cgc agc gcc gta gag atc gct ccc act cag gtc aca	1741
Leu Pro Pro Ser Arg Ser Ala Val Glu Ile Ala Pro Thr Gln Val Thr	
545 550 555	
gag act gat gag tgc cga ctg aac cag aac atc tgt ggc cac gga gag	1789
Glu Thr Asp Glu Cys Arg Leu Asn Gln Asn Ile Cys Gly His Gly Glu	
560 565 570 575	
tgc gtg ccg ggc ccc cct gac tac tcc tgc cac tgc aac ccc ggc tac	1837
Cys Val Pro Gly Pro Pro Asp Tyr Ser Cys His Cys Asn Pro Gly Tyr	
580 585 590	
cgg tca cat ccc cag cac cgc tac tgc gtg gat gtg aac gag tgc gag	1885
Arg Ser His Pro Gln His Arg Tyr Cys Val Asp Val Asn Glu Cys Glu	
595 600 605	
gca gag ccc tgt ggc ccg ggg agg ggc atc tgc atg aac acc ggc ggc	1933
Ala Glu Pro Cys Gly Pro Gly Arg Gly Ile Cys Met Asn Thr Gly Gly	
610 615 620	
tcc tac aat tgc cac tgc aac cgc ggc tac cgc ctg cac gtg ggc gcc	1981
Ser Tyr Asn Cys His Cys Asn Arg Gly Tyr Arg Leu His Val Gly Ala	
625 630 635	
ggg ggg cgc tcg tgc gtg gac ctg aac gaa tgc gcc aag ccc cac ctg	2029
Gly Gly Arg Ser Cys Val Asp Leu Asn Glu Cys Ala Lys Pro His Leu	
640 645 650 655	
tgc ggc gac ggc ggc ttc tgc atc aac ttt ccc ggt cac tac aag tgc	2077
Cys Gly Asp Gly Gly Phe Cys Ile Asn Phe Pro Gly His Tyr Lys Cys	
660 665 670	
aac tgc tac ccc ggc tac cgg ctc aaa gcc tcc cgg cct cct gtg tgc	2125
Asn Cys Tyr Pro Gly Tyr Arg Leu Lys Ala Ser Arg Pro Pro Val Cys	
675 680 685	

gaa gac atc gac gag tgc cgg gac cca agc tct tgc ccg gat ggc aaa	2173
Glu Asp Ile Asp Glu Cys Arg Asp Pro Ser Ser Cys Pro Asp Gly Lys	
690 695 700	
tgc gag aac aag ccc ggg agc ttc aag tgc atc gcc tgt cag cct ggc	2221
Cys Glu Asn Lys Pro Gly Ser Phe Lys Cys Ile Ala Cys Gln Pro Gly	
705 710 715	
tac cgc agc cag ggg ggc ggg gcc tgt cgc gac gtg aac gag tgc gcc	2269
Tyr Arg Ser Gln Gly Gly Gly Ala Cys Arg Asp Val Asn Glu Cys Ala	
720 725 730 735	
gag ggc agc ccc tgc tcg cct ggc tgg tgc gag aac ctc ccg ggc tcc	2317
Glu Gly Ser Pro Cys Ser Pro Gly Trp Cys Glu Asn Leu Pro Gly Ser	
740 745 750	
ttc cgc tgc acc tgt gcc cag ggc tac gcg ccc gcg ccc gac ggc cgc	2365
Phe Arg Cys Thr Cys Ala Gln Gly Tyr Ala Pro Ala Pro Asp Gly Arg	
755 760 765	
agt tgc ttg gat gtg gac gag tgt gag gct ggg gac gtg tgt gac aat	2413
Ser Cys Leu Asp Val Asp Glu Cys Glu Ala Gly Asp Val Cys Asp Asn	
770 775 780	
ggc atc tgc agc aac acg cca gga tct ttc cag tgt cag tgc ctc tct	2461
Gly Ile Cys Ser Asn Thr Pro Gly Ser Phe Gln Cys Gln Cys Leu Ser	
785 790 795	
ggc tac cat ctg tcc agg gac cgg agc cac tgc gag gac att gat gag	2509
Gly Tyr His Leu Ser Arg Asp Arg Ser His Cys Glu Asp Ile Asp Glu	
800 805 810 815	
tgt gac ttc cct gca gcc tgc att ggg ggt gac tgc atc aat acc aat	2557
Cys Asp Phe Pro Ala Ala Cys Ile Gly Gly Asp Cys Ile Asn Thr Asn	
820 825 830	
ggc tcc tac aga tgt ctt tgc ccc cag ggg cat cgg ctg gtg ggt ggc	2605
Gly Ser Tyr Arg Cys Leu Cys Pro Gln Gly His Arg Leu Val Gly Gly	
835 840 845	
agg aaa tgc caa gac ata gat gag tgc agc cag gac ccg agc ctg tgc	2653
Arg Lys Cys Gln Asp Ile Asp Glu Cys Ser Gln Asp Pro Ser Leu Cys	
850 855 860	
ctt ccc cat ggg gcc tgc aag aac ctt cag ggc tcc tat gtg tgt gtc	2701
Leu Pro His Gly Ala Cys Lys Asn Leu Gln Gly Ser Tyr Val Cys Val	
865 870 875	
tgc gat gag ggc ttc act ccc acc cag gac cag cac ggt tgt gag gag	2749
Cys Asp Glu Gly Phe Thr Pro Thr Gln Asp Gln His Gly Cys Glu Glu	
880 885 890 895	
gtg gag cag ccc cac cac aag aag gag tgc tac ctg aac ttc gat gac	2797
Val Glu Gln Pro His His Lys Lys Glu Cys Tyr Leu Asn Phe Asp Asp	
900 905 910	

aca gtg ttc tgc gac agc gta ttg gcc acc aac gtg acc cag cag gag	2845
Thr Val Phe Cys Asp Ser Val Leu Ala Thr Asn Val Thr Gln Gln Glu	
915 920 925	
tgc tgc tgc tct ctg ggg gcc ggc tgg ggc gac cac tgc gaa atc tac	2893
Cys Cys Cys Ser Leu Gly Ala Gly Trp Gly Asp His Cys Glu Ile Tyr	
930 935 940	
ccc tgc cca gtc tac agc tca gcc gag ttc cac agc ctc tgc cca gac	2941
Pro Cys Pro Val Tyr Ser Ser Ala Glu Phe His Ser Leu Cys Pro Asp	
945 950 955	
gga aag ggc tac acc cag gac aac aac atc gtc aac tac ggc atc cca	2989
Gly Lys Gly Tyr Thr Gln Asp Asn Asn Ile Val Asn Tyr Gly Ile Pro	
960 965 970 975	
gcc cac cgt gac atc gac gag tgc atg ttg ttc ggg tcg gag att tgc	3037
Ala His Arg Asp Ile Asp Glu Cys Met Leu Phe Gly Ser Glu Ile Cys	
980 985 990	
aag gag ggc aag tgc gtg aac acg cag cct ggc tac gag tgc tac tgc	3085
Lys Glu Gly Lys Cys Val Asn Thr Gln Pro Gly Tyr Glu Cys Tyr Cys	
995 1000 1005	
aag cag ggc ttc tac tac gac ggg aac ctg ctg gaa tgc gtg gac gtg	3133
Lys Gln Gly Phe Tyr Tyr Asp Gly Asn Leu Leu Glu Cys Val Asp Val	
1010 1015 1020	
gac gag tgc ctg gac gag tcc aac tgc cgg aac gga gtg tgt gag aac	3181
Asp Glu Cys Leu Asp Glu Ser Asn Cys Arg Asn Gly Val Cys Glu Asn	
1025 1030 1035	
acg cgc ggc ggc tac cgc tgt gcc tgc acg ccc cct gcc gag tac agt	3229
Thr Arg Gly Gly Tyr Arg Cys Ala Cys Thr Pro Pro Ala Glu Tyr Ser	
1040 1045 1050 1055	
ccc gcg cag cgc cag tgc ctg agc ccg gaa gag atg gac gtg gac gag	3277
Pro Ala Gln Arg Gln Cys Leu Ser Pro Glu Glu Met Asp Val Asp Glu	
1060 1065 1070	
tgc cag gac ccg gca gcc tgc cgc cct ggc cgc tgc gtc aac ctg ccg	3325
Cys Gln Asp Pro Ala Ala Cys Arg Pro Gly Arg Cys Val Asn Leu Pro	
1075 1080 1085	
ggc tcc tac cgc tgc gag tgt cgc ccg ccc tgg gtg ccc ggg ccc tcc	3373
Gly Ser Tyr Arg Cys Glu Cys Arg Pro Pro Trp Val Pro Gly Pro Ser	
1090 1095 1100	
ggc cgc gat tgc cag ctc ccc gag agc ccg gcc gag cgt gcc ccg gag	3421
Gly Arg Asp Cys Gln Leu Pro Glu Ser Pro Ala Glu Arg Ala Pro Glu	
1105 1110 1115	
cgg cgc gac gtg tgc tgg agc cag cgc gga gag gac ggc atg tgc gct	3469
Arg Arg Asp Val Cys Trp Ser Gln Arg Gly Glu Asp Gly Met Cys Ala	
1120 1125 1130 1135	

ggc ccc ctg gcc ggg cct gcc ctc acc ttc gac gac tgc tgc tgc cgc 3517
 Gly Pro Leu Ala Gly Pro Ala Leu Thr Phe Asp Asp Cys Cys Cys Arg
 1140 1145 1150

cag ggc cgc ggc tgg ggc gcc caa tgc cga ccg tgc ccg ccg cgc ggc 3565
 Gln Gly Arg Gly Trp Gly Ala Gln Cys Arg Pro Cys Pro Pro Arg Gly
 1155 1160 1165

gcg ggg tcc cat tgc ccg aca tcg cag agc gag agc aat tcc ttc tgg 3613
 Ala Gly Ser His Cys Pro Thr Ser Gln Ser Glu Ser Asn Ser Phe Trp
 1170 1175 1180

gac aca agc ccc ctg ctg ttg ggg aag ccc cca aga gat gag gac agt 3661
 Asp Thr Ser Pro Leu Leu Gly Lys Pro Pro Arg Asp Glu Asp Ser
 1185 1190 1195

tca gag gag gat tca gac gag tgt cgc tgc gtg agt ggc cgc tgc gtg 3709
 Ser Glu Glu Asp Ser Asp Glu Cys Arg Cys Val Ser Gly Arg Cys Val
 1200 1205 1210 1215

ccg cgg ccg ggc ggc gcc gcg tgc gag tgt ccc ggc ggc ttc cag ctc 3757
 Pro Arg Pro Gly Gly Ala Ala Cys Glu Cys Pro Gly Gly Phe Gln Leu
 1220 1225 1230

gac gcc tcc cgc gcc cgc tgc gtg gat atc gac gag tgc cga gag ctg 3805
 Asp Ala Ser Arg Ala Arg Cys Val Asp Ile Asp Glu Cys Arg Glu Leu
 1235 1240 1245

aac cag cgc ggg ctg ctg tgc aag agc gag cgc tgc gtg aac acc agc 3853
 Asn Gln Arg Gly Leu Leu Cys Lys Ser Glu Arg Cys Val Asn Thr Ser
 1250 1255 1260

ggc tcc ttc cgc tgc gtc tgc aaa gcc ggc ttc gcg cgc agc cgc ccg 3901
 Gly Ser Phe Arg Cys Val Cys Lys Ala Gly Phe Ala Arg Ser Arg Pro
 1265 1270 1275

cac ggg gcc tgc gtt ccc cag cgc cgc cgc tgacgccgcc gacgccgcc 3951
 His Gly Ala Cys Val Pro Gln Arg Arg Arg
 1280 1285

tcggcccaga cctcggtgat cactgagggg tttccgcgag ctcggcctca cttctgcccc 4011
 gacttggtggc tcggaccag ggaccttcag ggcccgcaga ccctcccggc gccttgagac 4071
 ccgagggccc cctaccggcc cccctccccg gttagcgggc gggtgtaagg tctccggcgg 4131
 gcgctgcctg ccttctctcc agaggggtgtt tcctagaaac tgataaatca gatcgtgcct 4191
 ctttaaaaaa aaaaaaaaaa aaaagggcgg ccgc 4225

<210> 8
 <211> 1289
 <212> PRT
 <213> Homo sapiens

<400> 8
 Met Arg Gly Ala Gly Ala Ala Gly Leu Leu Ala Leu Leu Leu Leu Leu
 1 5 10 15
 Leu Leu Leu Leu Leu Gly Leu Gly Gly Arg Val Glu Gly Gly Pro Ala
 20 25 30
 Gly Glu Arg Gly Ala Gly Gly Gly Gly Ala Leu Ala Arg Glu Arg Phe
 35 40 45

Lys Val Val Phe Ala Pro Val Ile Cys Lys Arg Thr Cys Leu Lys Gly
 50 55 60
 Gln Cys Arg Asp Ser Cys Gln Gln Gly Ser Asn Met Thr Leu Ile Gly
 65 70 75 80
 Glu Asn Gly His Ser Thr Asp Thr Leu Thr Gly Ser Gly Phe Arg Val
 85 90 95
 Val Val Cys Pro Leu Pro Cys Met Asn Gly Gly Gln Cys Ser Ser Arg
 100 105 110
 Asn Gln Cys Leu Cys Pro Pro Asp Phe Thr Gly Arg Phe Cys Gln Val
 115 120 125
 Pro Ala Gly Gly Ala Gly Gly Thr Gly Gly Ser Gly Pro Gly Leu
 130 135 140
 Ser Arg Thr Gly Ala Leu Ser Thr Gly Ala Leu Pro Pro Leu Ala Pro
 145 150 155 160
 Glu Gly Asp Ser Val Ala Ser Lys His Ala Ile Tyr Ala Val Gln Val
 165 170 175
 Ile Ala Asp Pro Pro Gly Pro Gly Glu Gly Pro Pro Ala Gln His Ala
 180 185 190
 Ala Phe Leu Val Pro Leu Gly Pro Gly Gln Ile Ser Ala Glu Val Gln
 195 200 205
 Ala Pro Pro Pro Val Val Asn Val Arg Val His His Pro Pro Glu Ala
 210 215 220
 Ser Val Gln Val His Arg Ile Glu Ser Ser Asn Ala Glu Ser Ala Ala
 225 230 235 240
 Pro Ser Gln His Leu Leu Pro His Pro Lys Pro Ser His Pro Arg Pro
 245 250 255
 Pro Thr Gln Lys Pro Leu Gly Arg Cys Phe Gln Asp Thr Leu Pro Lys
 260 265 270
 Gln Pro Cys Gly Ser Asn Pro Leu Pro Gly Leu Thr Lys Gln Glu Asp
 275 280 285
 Cys Cys Gly Ser Ile Gly Thr Ala Trp Gly Gln Ser Lys Cys His Lys
 290 295 300
 Cys Pro Gln Leu Gln Tyr Thr Gly Val Gln Lys Pro Gly Pro Val Arg
 305 310 315 320
 Gly Glu Val Gly Ala Asp Cys Pro Gln Gly Tyr Lys Arg Leu Asn Ser
 325 330 335
 Thr His Cys Gln Asp Ile Asn Glu Cys Ala Met Pro Gly Val Cys Arg
 340 345 350
 His Gly Asp Cys Leu Asn Asn Pro Gly Ser Tyr Arg Cys Val Cys Pro
 355 360 365
 Pro Gly His Ser Leu Gly Pro Ser Arg Thr Gln Cys Ile Ala Asp Lys
 370 375 380
 Pro Glu Glu Lys Ser Leu Cys Phe Arg Leu Val Ser Pro Glu His Gln
 385 390 395 400
 Cys Gln His Pro Leu Thr Thr Arg Leu Thr Arg Gln Leu Cys Cys Cys
 405 410 415
 Ser Val Gly Lys Ala Trp Gly Ala Arg Cys Gln Arg Cys Pro Thr Asp
 420 425 430
 Gly Thr Ala Ala Phe Lys Glu Ile Cys Pro Ala Gly Lys Gly Tyr His
 435 440 445
 Ile Leu Thr Ser His Gln Thr Leu Thr Ile Gln Gly Glu Ser Asp Phe
 450 455 460
 Ser Leu Phe Leu His Pro Asp Gly Pro Pro Lys Pro Gln Gln Leu Pro
 465 470 475 480
 Glu Ser Pro Ser Gln Ala Pro Pro Pro Glu Asp Thr Glu Glu Glu Arg
 485 490 495
 Gly Val Thr Thr Asp Ser Pro Val Ser Glu Glu Arg Ser Val Gln Gln
 500 505 510

Ser His Pro Thr Ala Thr Thr Thr Pro Ala Arg Pro Tyr Pro Glu Leu
 515 520 525
 Ile Ser Arg Pro Ser Pro Pro Thr Met Arg Trp Phe Leu Pro Asp Leu
 530 535 540
 Pro Pro Ser Arg Ser Ala Val Glu Ile Ala Pro Thr Gln Val Thr Glu
 545 550 555 560
 Thr Asp Glu Cys Arg Leu Asn Gln Asn Ile Cys Gly His Gly Glu Cys
 565 570 575
 Val Pro Gly Pro Pro Asp Tyr Ser Cys His Cys Asn Pro Gly Tyr Arg
 580 585 590
 Ser His Pro Gln His Arg Tyr Cys Val Asp Val Asn Glu Cys Glu Ala
 595 600 605
 Glu Pro Cys Gly Pro Gly Arg Gly Ile Cys Met Asn Thr Gly Gly Ser
 610 615 620
 Tyr Asn Cys His Cys Asn Arg Gly Tyr Arg Leu His Val Gly Ala Gly
 625 630 635 640
 Gly Arg Ser Cys Val Asp Leu Asn Glu Cys Ala Lys Pro His Leu Cys
 645 650 655
 Gly Asp Gly Gly Phe Cys Ile Asn Phe Pro Gly His Tyr Lys Cys Asn
 660 665 670
 Cys Tyr Pro Gly Tyr Arg Leu Lys Ala Ser Arg Pro Pro Val Cys Glu
 675 680 685
 Asp Ile Asp Glu Cys Arg Asp Pro Ser Ser Cys Pro Asp Gly Lys Cys
 690 695 700
 Glu Asn Lys Pro Gly Ser Phe Lys Cys Ile Ala Cys Gln Pro Gly Tyr
 705 710 715 720
 Arg Ser Gln Gly Gly Gly Ala Cys Arg Asp Val Asn Glu Cys Ala Glu
 725 730 735
 Gly Ser Pro Cys Ser Pro Gly Trp Cys Glu Asn Leu Pro Gly Ser Phe
 740 745 750
 Arg Cys Thr Cys Ala Gln Gly Tyr Ala Pro Ala Pro Asp Gly Arg Ser
 755 760 765
 Cys Leu Asp Val Asp Glu Cys Glu Ala Gly Asp Val Cys Asp Asn Gly
 770 775 780
 Ile Cys Ser Asn Thr Pro Gly Ser Phe Gln Cys Gln Cys Leu Ser Gly
 785 790 795 800
 Tyr His Leu Ser Arg Asp Arg Ser His Cys Glu Asp Ile Asp Glu Cys
 805 810 815
 Asp Phe Pro Ala Ala Cys Ile Gly Gly Asp Cys Ile Asn Thr Asn Gly
 820 825 830
 Ser Tyr Arg Cys Leu Cys Pro Gln Gly His Arg Leu Val Gly Gly Arg
 835 840 845
 Lys Cys Gln Asp Ile Asp Glu Cys Ser Gln Asp Pro Ser Leu Cys Leu
 850 855 860
 Pro His Gly Ala Cys Lys Asn Leu Gln Gly Ser Tyr Val Cys Val Cys
 865 870 875 880
 Asp Glu Gly Phe Thr Pro Thr Gln Asp Gln His Gly Cys Glu Glu Val
 885 890 895
 Glu Gln Pro His Lys Lys Glu Cys Tyr Leu Asn Phe Asp Asp Thr
 900 905 910
 Val Phe Cys Asp Ser Val Leu Ala Thr Asn Val Thr Gln Gln Glu Cys
 915 920 925
 Cys Cys Ser Leu Gly Ala Gly Trp Gly Asp His Cys Glu Ile Tyr Pro
 930 935 940
 Cys Pro Val Tyr Ser Ser Ala Glu Phe His Ser Leu Cys Pro Asp Gly
 945 950 955 960
 Lys Gly Tyr Thr Gln Asp Asn Asn Ile Val Asn Tyr Gly Ile Pro Ala
 965 970 975

His Arg Asp Ile Asp Glu Cys Met Leu Phe Gly Ser Glu Ile Cys Lys
 980 985 990
 Glu Gly Lys Cys Val Asn Thr Gln Pro Gly Tyr Glu Cys Tyr Cys Lys
 995 1000 1005
 Gln Gly Phe Tyr Tyr Asp Gly Asn Leu Leu Glu Cys Val Asp Val Asp
 1010 1015 1020
 Glu Cys Leu Asp Glu Ser Asn Cys Arg Asn Gly Val Cys Glu Asn Thr
 1025 1030 1035 1040
 Arg Gly Gly Tyr Arg Cys Ala Cys Thr Pro Pro Ala Glu Tyr Ser Pro
 1045 1050 1055
 Ala Gln Arg Gln Cys Leu Ser Pro Glu Glu Met Asp Val Asp Glu Cys
 1060 1065 1070
 Gln Asp Pro Ala Ala Cys Arg Pro Gly Arg Cys Val Asn Leu Pro Gly
 1075 1080 1085
 Ser Tyr Arg Cys Glu Cys Arg Pro Pro Trp Val Pro Gly Pro Ser Gly
 1090 1095 1100
 Arg Asp Cys Gln Leu Pro Glu Ser Pro Ala Glu Arg Ala Pro Glu Arg
 1105 1110 1115 1120
 Arg Asp Val Cys Trp Ser Gln Arg Gly Glu Asp Gly Met Cys Ala Gly
 1125 1130 1135
 Pro Leu Ala Gly Pro Ala Leu Thr Phe Asp Asp Cys Cys Cys Arg Gln
 1140 1145 1150
 Gly Arg Gly Trp Gly Ala Gln Cys Arg Pro Cys Pro Pro Arg Gly Ala
 1155 1160 1165
 Gly Ser His Cys Pro Thr Ser Gln Ser Glu Ser Asn Ser Phe Trp Asp
 1170 1175 1180
 Thr Ser Pro Leu Leu Leu Gly Lys Pro Pro Arg Asp Glu Asp Ser Ser
 1185 1190 1195 1200
 Glu Glu Asp Ser Asp Glu Cys Arg Cys Val Ser Gly Arg Cys Val Pro
 1205 1210 1215
 Arg Pro Gly Gly Ala Ala Cys Glu Cys Pro Gly Gly Phe Gln Leu Asp
 1220 1225 1230
 Ala Ser Arg Ala Arg Cys Val Asp Ile Asp Glu Cys Arg Glu Leu Asn
 1235 1240 1245
 Gln Arg Gly Leu Leu Cys Lys Ser Glu Arg Cys Val Asn Thr Ser Gly
 1250 1255 1260
 Ser Phe Arg Cys Val Cys Lys Ala Gly Phe Ala Arg Ser Arg Pro His
 1265 1270 1275 1280
 Gly Ala Cys Val Pro Gln Arg Arg Arg
 1285

<210> 9
 <211> 3867
 <212> DNA
 <213> Homo sapiens

<400> 9
 atgcgcgggg cgggggcggc ggggctgctg gcgctgctgc tgctgctgct gctgctgctg 60
 ctgggccttg gcggcagggg cgaggggggg ccggccggcg agcggggcgc aggcgggggc 120
 ggggcgctgg ccgcgagcgc cttcaaggtg gtcttttgcgc cggatgatctg caagcggacc 180
 tgtctcaagg gccagtgtcg ggacagttgt cagcaggggt ccaacatgac gctcatcgga 240
 gagaacggcc acagcacaga cacgctcacg ggctccggct tccgcgtggt ggtgtgccct 300
 ctccccctgca tgaatggcgg ccagtgtctc tcgcgaaacc agtgccctgtg tcccccgac 360
 ttactggggc gcttctgcca ggtgcccgcg ggaggagccg gtgggggtac cggcggtcca 420
 ggccccggcc tgcagcggac aggggcccctg tccacagggg cgtgcccgc cctggctccg 480
 gagggcgact ctgtggccag caagcacgcc atctacgccg tccaggtgat cgtgaccct 540
 cctgggcccg gggaggggcc tctgcccag cagcagcct tctgggtgcc cctaggcccg 600
 ggacagatct cagcagaagt gcaggccccg cccccgtgg tgaatgtgcg cgtccatcac 660

ccgccccgagg	cctcagtgcca	ggtgcaccgc	attgagagct	cgaacgccga	gagcgcagcc	720
ccctcccagc	acctgctgcc	gcaccccaag	ccctcgcacc	ctcggccgcc	caccacagaag	780
cccctggggc	gctgctttca	ggacactctg	cccaagcagc	cgtgtggcag	caacccccctc	840
cccggcctca	ccaagcagga	agactgctgc	ggtagcatcg	gcactgcctg	gggcccagagc	900
aagtgccaca	agtgtcccca	gctgcagtac	acaggagtgc	agaagccagg	gcctgtacgt	960
ggggaagtgg	gcgctgactg	tccccagggc	tacaagaggg	ttaacagcac	ccactgccag	1020
gacatcaacg	agtgcgcaat	gccggggcgtg	tgtcgccatg	gtgactgcct	caacaacccct	1080
ggctcctatc	gctgtgtctg	cccacctggc	catagtttag	gccccctccc	tacacagtgc	1140
attgcagaca	aaccggagga	gaagagcctg	tgtttccgcc	tggtgagccc	tgagcaccag	1200
tgccagcacc	cactgaccac	ccgctgacc	cgccagctct	gctgctgcag	tgtcggcaag	1260
gcctggggcg	cgcggtgtca	gcgctgcccc	acagatggca	ccgctgcgtt	caaggagatc	1320
tgcccagctg	ggaagggata	ccacattctc	acctcccacc	agacgctcac	cattcagggc	1380
gagagtgaat	tttccctttt	cctgcaccct	gacggggccac	ccaagcccca	gcagcttccg	1440
gagagcccta	gccaggctcc	accacctgag	gacacagagg	aagagagagg	ggtgaccacg	1500
gactcaccgg	tgagtgcgga	gaggtcagtg	cagcagagcc	acccaactgc	caccacgact	1560
cctgcccggc	cctaccccga	gctgatctcc	cgtccctcgc	ccccgaccat	gcgctggttc	1620
ctgcccggact	tgctccttcc	ccgcagcgcc	gtagagatcg	ctcccactca	ggtcacagag	1680
actgatgagt	gccgactgaa	ccagaacatc	tgtggccacg	gagagtgcgt	gccgggcccc	1740
cctgactact	cctgccactg	caacccccgg	taccggtcac	atccccagca	ccgctactgc	1800
gtggatgtga	acgagtgcga	ggcagagccc	tgtggccccg	ggagggggcat	ctgcatgaac	1860
accggcggct	cctacaattg	ccactgcaac	cgcggtacc	gcctgcacgt	gggcgcccggg	1920
gggcgctcgt	gcgtggacct	gaacgaatgc	gccaagcccc	acctgtgcgg	cgacggcggc	1980
ttctgcatca	actttcccgg	tactacaag	tgcaactgct	accccggtca	ccggctcaaa	2040
gcctcccggc	ctcctgtgtg	cgaagacatc	gacgagtgcc	gggacccaag	ctcttgcccg	2100
gatggcaaat	gcgagaacaa	gcccggggagc	ttcaagtgc	tcgcctgtca	gcctgggtac	2160
cgagccagg	ggggcggggc	ctgtcgcgac	gtgaacgagt	gcgccgaggg	cagcccctgc	2220
tcgcctggct	ggtgcgagaa	cctcccgggc	tccttccgct	gcacctgtgc	ccagggtac	2280
gcgcccgcgc	ccgacggccg	cagttgcttg	gatgtggacg	agtgtgaggc	tggggacgtg	2340
tgtgacaatg	gcactctgag	caacacgcca	ggatctttcc	agtgtcagtg	cctctctggc	2400
taccatctgt	ccagggaccg	gagccactgc	gaggacattg	atgagtgtga	cttccctgca	2460
gcctgcattg	ggggtgactg	catcaatacc	aatggctcct	acagatgtct	ttgccccag	2520
gggcacccgc	tgggtgggtg	caggaaatgc	caagacatag	atgagtgcag	ccaggacccg	2580
agcctgtgcc	ttccccatgg	ggcctgcaag	aaccttcagg	gtcctatgt	gtgtgtctgc	2640
gatgagggct	tcactcccac	ccaggaccag	cacggttgtg	aggaggtgga	gcagccccac	2700
cacaagaagg	agtgtctact	gaacttcgat	gacacagtgt	tctgcgacag	cgtattggcc	2760
accaacgtga	cccagcagga	gtgctgctgc	tctctggggg	ccggctgggg	cgaccactgc	2820
gaaatctacc	cctgcccagt	ctacagctca	gccgagttcc	acagcctctg	cccagacgga	2880
aagggttaca	cccaggacaa	caacatcgte	aactacggca	tcccagccca	ccgtgacatc	2940
gacgagtga	tgttgttcgg	gtcggagatt	tgcaaggagg	gcaagtgcgt	gaacacgcag	3000
cctggctacg	agtgtacttg	caagcagggc	ttctactacg	acgggaacct	gctggaatgc	3060
gtggacgtgg	acgagtgcct	ggacgagtc	aactgccgga	acggagtgtg	tgagaacacg	3120
cgcgccggct	accgctgtgc	ctgcacgccc	cctgccgagt	acagtcccgc	gcagcgccag	3180
tgcttgagcc	cgggaagagat	ggacgtggac	gagtgccagg	acccggcagc	ctgccgccct	3240
ggccgctgcg	tcaacctgcc	gggctcctac	cgctgcgagt	gtcgcccgc	ctgggtgccc	3300
gggcccctcc	gccgcgattg	ccagctcccc	gagagcccgg	ccgagcgtgc	cccggagcgg	3360
cgcgacgtgt	gctggagcca	gcgcggagag	gacggcatgt	gcgctggccc	cctggccggg	3420
cctgcccctca	ccttcgacga	ctgctgctgc	cgccagggcc	gcggctgggg	cgcccaatgc	3480
cgaccgtgcc	cgccgcgcgg	cgcggggtcc	cattgcccga	catcgagag	cgagagcaat	3540
tccttctggg	acacaagccc	cctgctgttg	gggaagcccc	caagagatga	ggacagttca	3600
gaggaggatt	cagacgagtg	tcgctgcgtg	agtggccgct	gcgtgccgcg	gccgggcggc	3660
gccgcgtgcg	agtgtcccgg	cggttccag	ctcgacgct	cccgcgcccg	ctgctgggat	3720
atcgacgagt	gccgagagct	gaaccagcgc	gggctgctgt	gcaagagcga	gcgctgcgtg	3780
aacaccagcg	gtccttccg	ctgctctgc	aaagccggct	tcgcgcgcag	ccgcccgcac	3840
ggggcctgcg	ttccccagcg	ccgcgcgc				3867

<210> 10

<211> 4316

<212> DNA

<213> Mus musculus

<220>

<221> CDS

<222> (157)...(3915)

<400> 10

cctcctgctg	tccccctccct	acccttggt	tctcgccccg	ctctgccctc	tgctaccaac	60
actcgatccc	ctgctcgggc	tcgacctcca	atctccgagg	gtcgtgcggc	cccggatgcc	120
cgggccccga	gcggtgcccc	cggcctggcc	cctgcg	atg	cgc cag gcc ggc gga	174
			Met	Arg	Gln Ala Gly Gly	
			1		5	
ttg ggg ctg ctg gca cta ctc ctg ctg gcg ctg ctg ggc ccc ggc ggc	222					
Leu Gly Leu Leu Ala Leu Leu Leu Leu Ala Leu Leu Gly Pro Gly Gly						
10 15 20						
cga ggg gtg ggc cgg ccg ggc agc ggg gca cag gcg ggg gcg ggg cgc	270					
Arg Gly Val Gly Arg Pro Gly Ser Gly Ala Gln Ala Gly Ala Gly Arg						
25 30 35						
tgg gcc caa cgc ttc aag gtg gtc ttt gcg cct gtg atc tgc aag cgg	318					
Trp Ala Gln Arg Phe Lys Val Val Phe Ala Pro Val Ile Cys Lys Arg						
40 45 50						
acc tgt ctg aag ggc cag tgt cgg gac agc tgt cag cag ggc tcc aac	366					
Thr Cys Leu Lys Gly Gln Cys Arg Asp Ser Cys Gln Gln Gly Ser Asn						
55 60 65 70						
atg acg ctc atc gga gag aac ggc cac agc acc gac acg ctc acc ggt	414					
Met Thr Leu Ile Gly Glu Asn Gly His Ser Thr Asp Thr Leu Thr Gly						
75 80 85						
tct gcc ttc cgc gtg gtg gtg tgc cct cta ccc tgc atg aac ggt ggc	462					
Ser Ala Phe Arg Val Val Val Cys Pro Leu Pro Cys Met Asn Gly Gly						
90 95 100						
cag tgc tct tcc cga aac cag tgc ctg tgt ccc ccg gat ttc acg ggg	510					
Gln Cys Ser Ser Arg Asn Gln Cys Leu Cys Pro Pro Asp Phe Thr Gly						
105 110 115						
cgc ttc tgc cag gtg cct gct gca gga acc gga gct ggc acc ggg agt	558					
Arg Phe Cys Gln Val Pro Ala Ala Gly Thr Gly Ala Gly Thr Gly Ser						
120 125 130						
tca ggc ccc ggc tgg ccc gac cgg gcc atg tcc aca ggc ccg ctg ccg	606					
Ser Gly Pro Gly Trp Pro Asp Arg Ala Met Ser Thr Gly Pro Leu Pro						
135 140 145 150						
ccc ctt gcc cca gaa gga gag tct gtg gct agc aaa cac gcc att tac	654					
Pro Leu Ala Pro Glu Gly Glu Ser Val Ala Ser Lys His Ala Ile Tyr						
155 160 165						
gcg gtg cag gtg atc gca gat cct ccc ggg ccg ggg gag ggt cct cct	702					
Ala Val Gln Val Ile Ala Asp Pro Pro Gly Pro Gly Glu Gly Pro Pro						
170 175 180						

gca caa cat gca gcc ttc ttg gtg ccc ctg ggg cca gga caa atc tcg Ala Gln His Ala Ala Phe Leu Val Pro Leu Gly Pro Gly Gln Ile Ser 185 190 195	750
gca gaa gtg cag gct ccg ccc ccc gtg gtg aac gtg cgt gtc cat cac Ala Glu Val Gln Ala Pro Pro Pro Val Val Asn Val Arg Val His His 200 205 210	798
cct cct gaa gct tcc gtt cag gtg cac cgc atc gag ggg ccg aac gct Pro Pro Glu Ala Ser Val Gln Val His Arg Ile Glu Gly Pro Asn Ala 215 220 225 230	846
gaa ggc cca gcc tct tcc cag cac ttg ctg ccg cat ccc aag ccc ccg Glu Gly Pro Ala Ser Ser Gln His Leu Leu Pro His Pro Lys Pro Pro 235 240 245	894
cac ccg agg cca ccc act caa aag cca ctg ggc cgc tgc ttc cag gac His Pro Arg Pro Pro Thr Gln Lys Pro Leu Gly Arg Cys Phe Gln Asp 250 255 260	942
aca ttg ccc aag cag cct tgt ggc agc aac cct ttg cct ggc ctt acc Thr Leu Pro Lys Gln Pro Cys Gly Ser Asn Pro Leu Pro Gly Leu Thr 265 270 275	990
aag cag gaa gat tgc tgc ggt agc atc ggt act gcc tgg gga caa agc Lys Gln Glu Asp Cys Cys Gly Ser Ile Gly Thr Ala Trp Gly Gln Ser 280 285 290	1038
aag tgt cac aag tgc cca cag ctt cag tat aca ggg gtg cag aag cct Lys Cys His Lys Cys Pro Gln Leu Gln Tyr Thr Gly Val Gln Lys Pro 295 300 305 310	1086
gta cct gta cgt ggg gag gtg ggt gct gac tgc ccc cag ggc tac aag Val Pro Val Arg Gly Glu Val Gly Ala Asp Cys Pro Gln Gly Tyr Lys 315 320 325	1134
agg ctc aac agc acc cac tgc cag gat atc aac gaa tgt gcg atg ccc Arg Leu Asn Ser Thr His Cys Gln Asp Ile Asn Glu Cys Ala Met Pro 330 335 340	1182
ggg aat gtg tgc cat ggt gac tgc ctc aac aac cct ggc tct tat cgc Gly Asn Val Cys His Gly Asp Cys Leu Asn Asn Pro Gly Ser Tyr Arg 345 350 355	1230
tgt gtc tgc ccg ccc ggt cat agc ttg ggt ccc ctc gca gca cag tgc Cys Val Cys Pro Pro Gly His Ser Leu Gly Pro Leu Ala Ala Gln Cys 360 365 370	1278
att gcc gac aaa cca gag gag aag agc ctg tgt ttc cgc ctt gtg agc Ile Ala Asp Lys Pro Glu Glu Lys Ser Leu Cys Phe Arg Leu Val Ser 375 380 385 390	1326
acc gaa cac cag tgc cag cac cct ctg acc aca cgc cta acc cgc cag Thr Glu His Gln Cys Gln His Pro Leu Thr Thr Arg Leu Thr Arg Gln 395 400 405	1374

ctc tgc tgc tgt agt gtg ggt aaa gcc tgg ggt gcc cgg tgc cag cgc	1422
Leu Cys Cys Cys Ser Val Gly Lys Ala Trp Gly Ala Arg Cys Gln Arg	
410 415 420	
tgc ccg gca gat ggt aca gca gcc ttc aag gag atc tgc ccc ggc tgg	1470
Cys Pro Ala Asp Gly Thr Ala Ala Phe Lys Glu Ile Cys Pro Gly Trp	
425 430 435	
gaa agg gta cca tat cct cac ctc cca cca gac gct cac cat cca ggg	1518
Glu Arg Val Pro Tyr Pro His Leu Pro Pro Asp Ala His His Pro Gly	
440 445 450	
gga aag cga ctt ctc cct ctt cct gca ccc gac ggg cca ccc aaa ccc	1566
Gly Lys Arg Leu Leu Pro Leu Pro Ala Pro Asp Gly Pro Pro Lys Pro	
455 460 465 470	
cag cag ctt cct gaa agc ccc agc cga gca cca ccc ctc gag gac aca	1614
Gln Gln Leu Pro Glu Ser Pro Ser Arg Ala Pro Pro Leu Glu Asp Thr	
475 480 485	
gag gaa gag aga gga gtg acc atg gat cca cca gtg agt gag gag cga	1662
Glu Glu Glu Arg Gly Val Thr Met Asp Pro Pro Val Ser Glu Glu Arg	
490 495 500	
tcg gtg cag cag agc cac ccc act acc acc acc tca ccc ccc cgg cct	1710
Ser Val Gln Gln Ser His Pro Thr Thr Thr Ser Pro Pro Arg Pro	
505 510 515	
tac cca gag ctc atc tct cgc ccc tcc cca cct acc ttc cac cgg ttc	1758
Tyr Pro Glu Leu Ile Ser Arg Pro Ser Pro Pro Thr Phe His Arg Phe	
520 525 530	
ctg cca gac ttg ccc cca tcc cga agt gca gtg gag atc gcc ccc act	1806
Leu Pro Asp Leu Pro Pro Ser Arg Ser Ala Val Glu Ile Ala Pro Thr	
535 540 545 550	
cag gtc aca gag acc gat gag tgc cga ttg aac cag aat atc tgt ggc	1854
Gln Val Thr Glu Thr Asp Glu Cys Arg Leu Asn Gln Asn Ile Cys Gly	
555 560 565	
cat gga cag tgt gtg cct ggc ccc tcg gat tac tcc tgc cac tgc aac	1902
His Gly Gln Cys Val Pro Gly Pro Ser Asp Tyr Ser Cys His Cys Asn	
570 575 580	
gct ggc tac cgg tca cac ccg cag cac cgc tac tgt gtt gat gtg aac	1950
Ala Gly Tyr Arg Ser His Pro Gln His Arg Tyr Cys Val Asp Val Asn	
585 590 595	
gag tgc gag gca gag ccc tgc ggc ccc ggg aaa ggc atc tgt atg aac	1998
Glu Cys Glu Ala Glu Pro Cys Gly Pro Gly Lys Gly Ile Cys Met Asn	
600 605 610	
act ggt ggc tcc tac aat tgt cac tgc aac cga ggc tac cgc ctc cac	2046
Thr Gly Gly Ser Tyr Asn Cys His Cys Asn Arg Gly Tyr Arg Leu His	
615 620 625 630	

gtg ggt gca ggg ggc cgc tcg tgc gtg gac ctg aac gag tgc gcc aag Val Gly Ala Gly Gly Arg Ser Cys Val Asp Leu Asn Glu Cys Ala Lys 635 640 645	2094
cct cac ctg tgt ggg gac ggt ggc ttc tgc atc aac ttc cct ggt cac Pro His Leu Cys Gly Asp Gly Gly Phe Cys Ile Asn Phe Pro Gly His 650 655 660	2142
tac aaa tgc aac tgc tat cct ggc tac cgg ctc aag gcc tcc cga ccg Tyr Lys Cys Asn Cys Tyr Pro Gly Tyr Arg Leu Lys Ala Ser Arg Pro 665 670 675	2190
ccc att tgc gaa gac atc gac gag tgt cgc gac cct agc acc tgc cct Pro Ile Cys Glu Asp Ile Asp Glu Cys Arg Asp Pro Ser Thr Cys Pro 680 685 690	2238
gat ggc aaa tgt gaa aac aaa cct ggc agc ttc aag tgc atc gcc tgc Asp Gly Lys Cys Glu Asn Lys Pro Gly Ser Phe Lys Cys Ile Ala Cys 695 700 705 710	2286
cag cct ggc tac cgt agc cag ggg ggc ggg gcc tgt cgt gat gtc aac Gln Pro Gly Tyr Arg Ser Gln Gly Gly Gly Ala Cys Arg Asp Val Asn 715 720 725	2334
gaa tgc tcc gag ggt acc ccc tgc tct cct gga tgg tgt gag aac ctt Glu Cys Ser Glu Gly Thr Pro Cys Ser Pro Gly Trp Cys Glu Asn Leu 730 735 740	2382
ccg ggt tct tac cgt tgc acg tgt gcc cag ggg ata cga acc cgc aca Pro Gly Ser Tyr Arg Cys Thr Cys Ala Gln Gly Ile Arg Thr Arg Thr 745 750 755	2430
gga cgc ctc agt tgc ata gac gtg gat gag tgt gag gct ggg aaa gtg Gly Arg Leu Ser Cys Ile Asp Val Asp Glu Cys Glu Ala Gly Lys Val 760 765 770	2478
tgc caa gat ggc atc tgc acg aac aca cca ggc tct ttc cag tgt cag Cys Gln Asp Gly Ile Cys Thr Asn Thr Pro Gly Ser Phe Gln Cys Gln 775 780 785 790	2526
tgc ctc tcc ggc tat cat ctg tca agg gat cgg agc cgc tgt gag gac Cys Leu Ser Gly Tyr His Leu Ser Arg Asp Arg Ser Arg Cys Glu Asp 795 800 805	2574
att gat gaa tgt gac ttc cct gcg gcc tgc atc ggg ggt gac tgc atc Ile Asp Glu Cys Asp Phe Pro Ala Ala Cys Ile Gly Gly Asp Cys Ile 810 815 820	2622
aat acc aat ggt tcc tac aga tgt ctc tgt ccc ctg ggt cat cgg ttg Asn Thr Asn Gly Ser Tyr Arg Cys Leu Cys Pro Leu Gly His Arg Leu 825 830 835	2670
gtg ggc ggc agg aag tgc aag aaa gat ata gat gag tgc agc cag gac Val Gly Gly Arg Lys Cys Lys Lys Asp Ile Asp Glu Cys Ser Gln Asp 840 845 850	2718

cca ggc ctg tgc ctg ccc cat gcc tgc gag aac ctc cag ggc tcc tat	2766
Pro Gly Leu Cys Leu Pro His Ala Cys Glu Asn Leu Gln Gly Ser Tyr	
855 860 865 870	
gtc tgt gtc tgt gat gag ggt ttc aca ctc acc cag gac cag cat ggg	2814
Val Cys Val Cys Asp Glu Gly Phe Thr Leu Thr Gln Asp Gln His Gly	
875 880 885	
tgt gag gag gtg gag cag ccc cac cac aag aag gag tgc tac ctt aac	2862
Cys Glu Glu Val Glu Gln Pro His His Lys Lys Glu Cys Tyr Leu Asn	
890 895 900	
ttc gat gac aca gtg ttc tgt gac agc gta ttg gct acc aat gtc act	2910
Phe Asp Asp Thr Val Phe Cys Asp Ser Val Leu Ala Thr Asn Val Thr	
905 910 915	
cag cag gaa tgc tgt tgc tct ctg gga gct ggc tgg gga gac cac tgc	2958
Gln Gln Glu Cys Cys Cys Ser Leu Gly Ala Gly Trp Gly Asp His Cys	
920 925 930	
gaa atc tat ccc tgt cca gtc tac agc tca gcc gaa ttt cac agc ctg	3006
Glu Ile Tyr Pro Cys Pro Val Tyr Ser Ser Ala Glu Phe His Ser Leu	
935 940 945 950	
gtg cct gat ggg aaa agg cta cac tca gga caa caa cat tgt gaa cta	3054
Val Pro Asp Gly Lys Arg Leu His Ser Gly Gln Gln His Cys Glu Leu	
955 960 965	
tgc att cct gcc cac cgt gac atc gac gaa tgc ata ttg ttt ggg gca	3102
Cys Ile Pro Ala His Arg Asp Ile Asp Glu Cys Ile Leu Phe Gly Ala	
970 975 980	
gag atc tgc aag gag ggc aag tgt gtg aac acg cag ccc ggc tac gag	3150
Glu Ile Cys Lys Glu Gly Lys Cys Val Asn Thr Gln Pro Gly Tyr Glu	
985 990 995	
tgc tac tgc aag cag ggc ttc tac tac gat ggc aac ctg ctg gag tgc	3198
Cys Tyr Cys Lys Gln Gly Phe Tyr Tyr Asp Gly Asn Leu Leu Glu Cys	
1000 1005 1010	
gtg gac gtg gat gag tgc ttg gat gag tct aac tgc agg aac gga gtg	3246
Val Asp Val Asp Glu Cys Leu Asp Glu Ser Asn Cys Arg Asn Gly Val	
1015 1020 1025 1030	
tgt gag aac aca cgt ggc ggc tac cgc tgt gcc tgc act ccg ccg gca	3294
Cys Glu Asn Thr Arg Gly Gly Tyr Arg Cys Ala Cys Thr Pro Pro Ala	
1035 1040 1045	
gag tac agt ccc gca cag gcc cag tgt ctg atc ccg gag aga tgg agc	3342
Glu Tyr Ser Pro Ala Gln Ala Gln Cys Leu Ile Pro Glu Arg Trp Ser	
1050 1055 1060	
acg ccc cag aga gac gtg aag tgt gct ggg gcc agc gag gag agg acg	3390
Thr Pro Gln Arg Asp Val Lys Cys Ala Gly Ala Ser Glu Glu Arg Thr	
1065 1070 1075	

gca tgt gta tgg ggc ccc tgg gcg gga cct gcc ctc act ttt gat gac	3438
Ala Cys Val Trp Gly Pro Trp Ala Gly Pro Ala Leu Thr Phe Asp Asp	
1080 1085 1090	
tgc tgc tgc cgc cag ccg cgg ctg ggt acc cag tgc aga ccg tgc ccg	3486
Cys Cys Cys Arg Gln Pro Arg Leu Gly Thr Gln Cys Arg Pro Cys Pro	
1095 1100 1105 1110	
cca cgt ggc acc ggg tcc cag tgc ccg act tca cag agt gag agc aat	3534
Pro Arg Gly Thr Gly Ser Gln Cys Pro Thr Ser Gln Ser Glu Ser Asn	
1115 1120 1125	
tct ttc tgg gac aca agc ccc ctg cta ctg ggg aag tct ccg cga gac	3582
Ser Phe Trp Asp Thr Ser Pro Leu Leu Gly Lys Ser Pro Arg Asp	
1130 1135 1140	
gaa gac agc tca gag gag gat tca gat gag tgc cgt tgt gtg agc gga	3630
Glu Asp Ser Ser Glu Glu Asp Ser Asp Glu Cys Arg Cys Val Ser Gly	
1145 1150 1155	
cgc tgt gtg cca cgg cca ggc ggg gcg gta tgc gag tgt cct gga ggc	3678
Arg Cys Val Pro Arg Pro Gly Gly Ala Val Cys Glu Cys Pro Gly Gly	
1160 1165 1170	
ttt cag ctg gac gcc tcc cgt gcc cgc tgc gtg gac att gat gag tgc	3726
Phe Gln Leu Asp Ala Ser Arg Ala Arg Cys Val Asp Ile Asp Glu Cys	
1175 1180 1185 1190	
cga gaa ctg aac cag cgg gga ctg ctg tgt aag agc gag cgg tgc gtg	3774
Arg Glu Leu Asn Gln Arg Gly Leu Leu Cys Lys Ser Glu Arg Cys Val	
1195 1200 1205	
aac acc agt gga tcc ttc cgc tgt gtc tgc aaa gct ggc ttc acg cgc	3822
Asn Thr Ser Gly Ser Phe Arg Cys Val Cys Lys Ala Gly Phe Thr Arg	
1210 1215 1220	
agc cgc cct cac ggg cct gcg tgc ctc agc gcc gcc gct gat gat gca	3870
Ser Arg Pro His Gly Pro Ala Cys Leu Ser Ala Ala Ala Asp Asp Ala	
1225 1230 1235	
gcc ata gcc cac acc tca gtg atc gat cat cga ggg tat ttt cac	3915
Ala Ile Ala His Thr Ser Val Ile Asp His Arg Gly Tyr Phe His	
1240 1245 1250	
tgaaagtgga gacagacaag tacatccttt gctcctgacc aaacgagagc atggacccaa	3975
ggatccttca gggcccacaa atctccttcc cacaccccaa acccaaggtg ctctgtctg	4035
cagagtgtg tctgctttct cccaaggggtg attcctagaa acttcgacat cagatctgcc	4095
cctttaattt actcttggtt tccaaggcaa attgatattc acatccaaaag cgggcagcat	4155
caactgcttg gcgggttgga ctgagctggg acccaggatg tgaaatagaa tttattgtgg	4215
ctctgattat gtacactaga tgtgcctgac ctgctgacca ggctcacatg gtttgtacaa	4275
taaatacatc cgccgggaaa aaaaaaaaaa aaaaaaaaaa a	4316

<210> 11

<211> 1253

<212> PRT

<213> Mus musculus

<400> 11

```

Met Arg Gln Ala Gly Gly Leu Gly Leu Leu Ala Leu Leu Leu Leu Ala
 1           5           10           15
Leu Leu Gly Pro Gly Gly Arg Gly Val Gly Arg Pro Gly Ser Gly Ala
           20           25           30
Gln Ala Gly Ala Gly Arg Trp Ala Gln Arg Phe Lys Val Val Phe Ala
           35           40           45
Pro Val Ile Cys Lys Arg Thr Cys Leu Lys Gly Gln Cys Arg Asp Ser
 50           55           60
Cys Gln Gln Gly Ser Asn Met Thr Leu Ile Gly Glu Asn Gly His Ser
65           70           75           80
Thr Asp Thr Leu Thr Gly Ser Ala Phe Arg Val Val Val Cys Pro Leu
           85           90           95
Pro Cys Met Asn Gly Gly Gln Cys Ser Ser Arg Asn Gln Cys Leu Cys
           100          105          110
Pro Pro Asp Phe Thr Gly Arg Phe Cys Gln Val Pro Ala Ala Gly Thr
           115          120          125
Gly Ala Gly Thr Gly Ser Ser Gly Pro Gly Trp Pro Asp Arg Ala Met
           130          135          140
Ser Thr Gly Pro Leu Pro Pro Leu Ala Pro Glu Gly Glu Ser Val Ala
145           150          155          160
Ser Lys His Ala Ile Tyr Ala Val Gln Val Ile Ala Asp Pro Pro Gly
           165          170          175
Pro Gly Glu Gly Pro Pro Ala Gln His Ala Ala Phe Leu Val Pro Leu
           180          185          190
Gly Pro Gly Gln Ile Ser Ala Glu Val Gln Ala Pro Pro Pro Val Val
           195          200          205
Asn Val Arg Val His His Pro Pro Glu Ala Ser Val Gln Val His Arg
           210          215          220
Ile Glu Gly Pro Asn Ala Glu Gly Pro Ala Ser Ser Gln His Leu Leu
225           230          235          240
Pro His Pro Lys Pro Pro His Pro Arg Pro Pro Thr Gln Lys Pro Leu
           245          250          255
Gly Arg Cys Phe Gln Asp Thr Leu Pro Lys Gln Pro Cys Gly Ser Asn
           260          265          270
Pro Leu Pro Gly Leu Thr Lys Gln Glu Asp Cys Cys Gly Ser Ile Gly
           275          280          285
Thr Ala Trp Gly Gln Ser Lys Cys His Lys Cys Pro Gln Leu Gln Tyr
           290          295          300
Thr Gly Val Gln Lys Pro Val Pro Val Arg Gly Glu Val Gly Ala Asp
305           310          315          320
Cys Pro Gln Gly Tyr Lys Arg Leu Asn Ser Thr His Cys Gln Asp Ile
           325          330          335
Asn Glu Cys Ala Met Pro Gly Asn Val Cys His Gly Asp Cys Leu Asn
           340          345          350
Asn Pro Gly Ser Tyr Arg Cys Val Cys Pro Pro Gly His Ser Leu Gly
           355          360          365
Pro Leu Ala Ala Gln Cys Ile Ala Asp Lys Pro Glu Glu Lys Ser Leu
           370          375          380
Cys Phe Arg Leu Val Ser Thr Glu His Gln Cys Gln His Pro Leu Thr
385           390          395          400
Thr Arg Leu Thr Arg Gln Leu Cys Cys Cys Ser Val Gly Lys Ala Trp
           405          410          415
Gly Ala Arg Cys Gln Arg Cys Pro Ala Asp Gly Thr Ala Ala Phe Lys
           420          425          430
Glu Ile Cys Pro Gly Trp Glu Arg Val Pro Tyr Pro His Leu Pro Pro
           435          440          445

```

Asp	Ala	His	His	Pro	Gly	Gly	Lys	Arg	Leu	Leu	Pro	Leu	Pro	Ala	Pro	
450						455					460					
Asp	Gly	Pro	Pro	Lys	Pro	Gln	Gln	Leu	Pro	Glu	Ser	Pro	Ser	Arg	Ala	
465					470					475					480	
Pro	Pro	Leu	Glu	Asp	Thr	Glu	Glu	Glu	Arg	Gly	Val	Thr	Met	Asp	Pro	
				485					490					495		
Pro	Val	Ser	Glu	Glu	Arg	Ser	Val	Gln	Gln	Ser	His	Pro	Thr	Thr	Thr	
			500					505					510			
Thr	Ser	Pro	Pro	Arg	Pro	Tyr	Pro	Glu	Leu	Ile	Ser	Arg	Pro	Ser	Pro	
		515					520					525				
Pro	Thr	Phe	His	Arg	Phe	Leu	Pro	Asp	Leu	Pro	Pro	Ser	Arg	Ser	Ala	
530					535					540						
Val	Glu	Ile	Ala	Pro	Thr	Gln	Val	Thr	Glu	Thr	Asp	Glu	Cys	Arg	Leu	
545					550					555					560	
Asn	Gln	Asn	Ile	Cys	Gly	His	Gly	Gln	Cys	Val	Pro	Gly	Pro	Ser	Asp	
			565						570					575		
Tyr	Ser	Cys	His	Cys	Asn	Ala	Gly	Tyr	Arg	Ser	His	Pro	Gln	His	Arg	
			580					585					590			
Tyr	Cys	Val	Asp	Val	Asn	Glu	Cys	Glu	Ala	Glu	Pro	Cys	Gly	Pro	Gly	
		595					600					605				
Lys	Gly	Ile	Cys	Met	Asn	Thr	Gly	Gly	Ser	Tyr	Asn	Cys	His	Cys	Asn	
610						615					620					
Arg	Gly	Tyr	Arg	Leu	His	Val	Gly	Ala	Gly	Gly	Arg	Ser	Cys	Val	Asp	
625					630					635					640	
Leu	Asn	Glu	Cys	Ala	Lys	Pro	His	Leu	Cys	Gly	Asp	Gly	Gly	Phe	Cys	
				645					650					655		
Ile	Asn	Phe	Pro	Gly	His	Tyr	Lys	Cys	Asn	Cys	Tyr	Pro	Gly	Tyr	Arg	
			660					665					670			
Leu	Lys	Ala	Ser	Arg	Pro	Pro	Ile	Cys	Glu	Asp	Ile	Asp	Glu	Cys	Arg	
		675					680					685				
Asp	Pro	Ser	Thr	Cys	Pro	Asp	Gly	Lys	Cys	Glu	Asn	Lys	Pro	Gly	Ser	
690						695					700					
Phe	Lys	Cys	Ile	Ala	Cys	Gln	Pro	Gly	Tyr	Arg	Ser	Gln	Gly	Gly	Gly	
705					710					715					720	
Ala	Cys	Arg	Asp	Val	Asn	Glu	Cys	Ser	Glu	Gly	Thr	Pro	Cys	Ser	Pro	
				725					730					735		
Gly	Trp	Cys	Glu	Asn	Leu	Pro	Gly	Ser	Tyr	Arg	Cys	Thr	Cys	Ala	Gln	
			740					745					750			
Gly	Ile	Arg	Thr	Arg	Thr	Gly	Arg	Leu	Ser	Cys	Ile	Asp	Val	Asp	Glu	
		755					760					765				
Cys	Glu	Ala	Gly	Lys	Val	Cys	Gln	Asp	Gly	Ile	Cys	Thr	Asn	Thr	Pro	
		770				775					780					
Gly	Ser	Phe	Gln	Cys	Gln	Cys	Leu	Ser	Gly	Tyr	His	Leu	Ser	Arg	Asp	
785					790					795					800	
Arg	Ser	Arg	Cys	Glu	Asp	Ile	Asp	Glu	Cys	Asp	Phe	Pro	Ala	Ala	Cys	
				805					810					815		
Ile	Gly	Gly	Asp	Cys	Ile	Asn	Thr	Asn	Gly	Ser	Tyr	Arg	Cys	Leu	Cys	
			820					825					830			
Pro	Leu	Gly	His	Arg	Leu	Val	Gly	Arg	Lys	Cys	Lys	Lys	Asp	Ile		
		835					840					845				
Asp	Glu	Cys	Ser	Gln	Asp	Pro	Gly	Leu	Cys	Leu	Pro	His	Ala	Cys	Glu	
850						855					860					
Asn	Leu	Gln	Gly	Ser	Tyr	Val	Cys	Val	Cys	Asp	Glu	Gly	Phe	Thr	Leu	
865					870					875					880	
Thr	Gln	Asp	Gln	His	Gly	Cys	Glu	Glu	Val	Glu	Gln	Pro	His	His	Lys	
				885					890					895		
Lys	Glu	Cys	Tyr	Leu	Asn	Phe	Asp	Asp	Thr	Val	Phe	Cys	Asp	Ser	Val	
			900					905					910			

Leu Ala Thr Asn Val Thr Gln Gln Glu Cys Cys Cys Ser Leu Gly Ala
 915 920 925
 Gly Trp Gly Asp His Cys Glu Ile Tyr Pro Cys Pro Val Tyr Ser Ser
 930 935 940
 Ala Glu Phe His Ser Leu Val Pro Asp Gly Lys Arg Leu His Ser Gly
 945 950 955 960
 Gln Gln His Cys Glu Leu Cys Ile Pro Ala His Arg Asp Ile Asp Glu
 965 970 975
 Cys Ile Leu Phe Gly Ala Glu Ile Cys Lys Glu Gly Lys Cys Val Asn
 980 985 990
 Thr Gln Pro Gly Tyr Glu Cys Tyr Cys Lys Gln Gly Phe Tyr Tyr Asp
 995 1000 1005
 Gly Asn Leu Leu Glu Cys Val Asp Val Asp Glu Cys Leu Asp Glu Ser
 1010 1015 1020
 Asn Cys Arg Asn Gly Val Cys Glu Asn Thr Arg Gly Gly Tyr Arg Cys
 1025 1030 1035 1040
 Ala Cys Thr Pro Pro Ala Glu Tyr Ser Pro Ala Gln Ala Gln Cys Leu
 1045 1050 1055
 Ile Pro Glu Arg Trp Ser Thr Pro Gln Arg Asp Val Lys Cys Ala Gly
 1060 1065 1070
 Ala Ser Glu Glu Arg Thr Ala Cys Val Trp Gly Pro Trp Ala Gly Pro
 1075 1080 1085
 Ala Leu Thr Phe Asp Asp Cys Cys Cys Arg Gln Pro Arg Leu Gly Thr
 1090 1095 1100
 Gln Cys Arg Pro Cys Pro Pro Arg Gly Thr Gly Ser Gln Cys Pro Thr
 1105 1110 1115 1120
 Ser Gln Ser Glu Ser Asn Ser Phe Trp Asp Thr Ser Pro Leu Leu Leu
 1125 1130 1135
 Gly Lys Ser Pro Arg Asp Glu Asp Ser Ser Glu Glu Asp Ser Asp Glu
 1140 1145 1150
 Cys Arg Cys Val Ser Gly Arg Cys Val Pro Arg Pro Gly Gly Ala Val
 1155 1160 1165
 Cys Glu Cys Pro Gly Gly Phe Gln Leu Asp Ala Ser Arg Ala Arg Cys
 1170 1175 1180
 Val Asp Ile Asp Glu Cys Arg Glu Leu Asn Gln Arg Gly Leu Leu Cys
 1185 1190 1195 1200
 Lys Ser Glu Arg Cys Val Asn Thr Ser Gly Ser Phe Arg Cys Val Cys
 1205 1210 1215
 Lys Ala Gly Phe Thr Arg Ser Arg Pro His Gly Pro Ala Cys Leu Ser
 1220 1225 1230
 Ala Ala Ala Asp Asp Ala Ala Ile Ala His Thr Ser Val Ile Asp His
 1235 1240 1245
 Arg Gly Tyr Phe His
 1250

<210> 12

<211> 3759

<212> DNA

<213> Mus musculus

<400> 12

atgcgccagg	ccggcggatt	ggggctgctg	gcactactcc	tgctggcgct	gctgggcccc	60
ggcgggccgag	gggtgggccc	gccggggcagc	ggggcacagg	cgggggcggg	gcgctgggccc	120
caacgcttca	aggtggtctt	tgcgctgtg	atctgcaagc	ggacctgtct	gaagggccag	180
tgctcgggaca	gctgtcagca	gggctccaac	atgacgtca	tcggagagaa	cggccacagc	240
accgacacgc	tcaccggttc	tgccttcgcg	gtggtggtgt	gccctctacc	ctgcatgaac	300
ggtggccagt	gctcttccc	aaaccagtgc	ctgtgtcccc	cggatttcac	ggggcgcttc	360
tgccagggtgc	ctgctgcagg	aaccggagct	ggcaccggga	gttcaggccc	cggctggccc	420

gaccgggcca	tgtccacagg	cccgtgccc	ccccttggcc	cagaaggaga	gtctgtggct	480
agcaaacacg	ccatttacgc	ggtgcagggtg	atcgcagatc	ctcccgggccc	gggggagggt	540
cctcctgcac	aacatgcagc	cttcttggtg	cccctggggc	caggacaaat	ctcggcagaa	600
gtgcaggctc	cgccccccgt	ggtgaacgtg	cgtgtccatc	accctcctga	agcttccgtt	660
caggtgcacc	gcacgcagg	gccgaacgct	gaaggcccag	cctcttccca	gcacttgctg	720
ccgcatccca	agcccccgca	cccagggcca	cccactcaaa	agccactggg	ccgctgcttc	780
caggacacat	tgcccaagca	gccttggtgc	agcaaccctt	tgccctggcct	taccaagcag	840
gaagattgct	gcggttagcat	cggtactgcc	tggggacaaa	gcaagtgtca	caagtgccca	900
cagcttcagt	atacaggggt	gcagaagcct	gtacctgtac	gtggggagggt	gggtgctgac	960
tgcccccagg	gctacaagag	gctcaacagc	accactgcc	aggatatcaa	cgaatgtgcg	1020
atgcccggga	atgtgtgcc	tgggtgactgc	ctcaacaacc	ctggctctta	tcgctgtgtc	1080
tgcccggccg	gtcatagctt	gggtcccctc	gcagcacagt	gcattgccga	caaaccagag	1140
gagaagagcc	tgtgtttccg	ccttggtgagc	accgaacacc	agtgccagca	ccctctgacc	1200
acacgcctaa	cccgccagct	ctgctgctgt	agtgtgggta	aagcctgggg	tgcccgggtg	1260
cagcgtgcc	cggcagatgg	tacagcagct	ttcaaggaga	tctgccccgg	ctgggaaagg	1320
gtaccatata	ctcacctccc	accagacgct	caccatccag	ggggaaagcg	acttctccct	1380
cttctgcac	ccgacggggc	acccaaaccc	cagcagcttc	ctgaaagccc	cagccgagca	1440
ccaccctcg	aggacacaga	ggaagagaga	ggagtgaacca	tggatccacc	agtgagttag	1500
gagcgatcgg	tgacgcagag	ccacccact	accaccacct	cacccccccg	gccttaccga	1560
gagctcatct	ctcgcctctc	cccacctacc	ttccaccggt	tcctgccaga	cttgccccca	1620
tcccgaaagt	cagtggagat	cgccccact	caggtcacag	agaccgatga	gtgccgattg	1680
aaccagaata	tctgtggcca	tggacagtgt	gtgcctggcc	cctcggatta	ctcctgccac	1740
tgcaacgctg	gctaccgggtc	acaccgcag	caccgctact	gtgttgatgt	gaacgagtgc	1800
gaggcagagc	cctgcggccc	cgggaaaggc	atctgtatga	acactggtgg	ctcctacaat	1860
tgtcactgca	accgaggcta	ccgcctccac	gtgggtgcag	ggggccgctc	gtgcgtggac	1920
ctgaacgagt	gcgccaagcc	tcacctgtgt	ggggacgggtg	gcttctgcat	caacttccct	1980
ggtcactaca	aatgcaactg	ctatcctggc	taccggctca	aggcctcccg	accgcccatt	2040
tgcaagaca	tcgacgagt	tcgcgaccct	agcacctgcc	ctgatggcaa	atgtgaaaac	2100
aaacctggca	gcttcaagt	catcgcctgc	cagcctgggt	accgtagcca	ggggggcggg	2160
gcctgtcgtg	atgtcaacga	atgtccgag	ggtacccct	gctctcctgg	atggtgtgag	2220
aaccttccgg	gttcttaccg	ttgcacgtgt	gcccagggga	tacgaaccgg	cacaggacgc	2280
ctcagttgca	tagacgtgga	tgagtgtgag	gctgggaaag	tgtgccaaaga	tggcatctgc	2340
acgaacacac	caggctcttt	ccagtgtcag	tgcctctccg	gctatcatct	gtcaagggat	2400
cggagccgct	gtgaggacat	tgatgaatgt	gacttccctg	cggcctgcat	cgggggtgac	2460
tgcatcaata	ccaatggttc	ctacagatgt	ctctgtcccc	tgggtcatcg	gttggtgggc	2520
ggcaggaagt	gcaagaaaga	tatagatgag	tgcagccagg	accagggcct	gtgcctggcc	2580
catgcctgcg	agaacctcca	gggtccttat	gtctgtgtct	gtgatgagg	tttcacactc	2640
accaggacc	agcatgggtg	tgaggagggtg	gagcagcccc	accacaagaa	ggagtgtctac	2700
cttaacttcg	atgacacagt	gttctgtgac	agcgtattgg	ctaccaatgt	cactcagcag	2760
gaatgctgtt	gctctctggg	agctggctgg	ggagaccact	gcgaaatcta	tcctgtcca	2820
gtctacagct	cagccgaatt	tcacagcctg	gtgcctgatg	ggaaaaggct	acactcagga	2880
caacaacatt	gtgaactatg	cattcctgcc	caccgtgaca	tcgacgaatg	catattgttt	2940
ggggcagaga	tctgcaagga	gggcaagtgt	gtgaacacgc	agcccggcta	cagagtgtac	3000
tgcaagcagg	gcttctacta	cgatggcaac	ctgctggagt	gcgtggacgt	ggatgagtgc	3060
ttggatgagt	ctaactgcag	gaacggagt	tgtgagaaca	cacgtggcgg	ctaccgctgt	3120
gcctgcactc	cgccggcaga	gtacagtccc	gcacaggccc	agtgtctgat	cccggagaga	3180
tggagcacgc	cccagagaga	cgtgaagtgt	gctggggcca	gcgaggagag	gacggcatgt	3240
gtatggggcc	cctgggcggg	acctgcccct	acttttgatg	actgctgctg	ccgccagcgc	3300
cggctgggta	cccagtgcag	accgtgcccg	ccacgtggca	ccgggtccca	gtgcccgact	3360
tcacagagt	agagcaattc	tttctgggac	acaagcccc	tgctactggg	gaagtctccg	3420
cgagacgaag	acagctcaga	ggaggattca	gatgagtgcc	gttgtgtgag	cggacgctgt	3480
gtgccacggc	caggcggggc	ggtatgcgag	tgtcctggag	gctttcagct	ggacgcctcc	3540
cgtgcccgtc	gcgtggacat	tgatgagtgc	cgagaactga	accagcgggg	actgctgtgt	3600
aagagcgagc	ggtgcgtgaa	caccagtggg	tccttccgct	gtgtctgcaa	agctggcttc	3660
acgcgcagcc	gccctcacgg	gctcgcgtg	ctcagcgcg	ccgctgatga	tgacagccata	3720
gcccacacct	cagtgatcga	tcacgcaggg	tattttcac			3759

<210> 13
 <211> 1332
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (31)...(1113)

<400> 13

```

ccaagaattc ggcacgagga gagggccggcc atg gcc agc ctg ggg ctg ctg ctc      54
                               Met Ala Ser Leu Gly Leu Leu Leu
                               1                               5

ctg ctg tta ctg aca gca ctg cca ccg ctg tgg tcc tcc tca ctg cct      102
Leu Leu Leu Leu Thr Ala Leu Pro Pro Leu Trp Ser Ser Ser Leu Pro
   10                               15                               20

ggg ctg gac act gct gaa agt aaa gcc acc att gca gac ctg atc ctg      150
Gly Leu Asp Thr Ala Glu Ser Lys Ala Thr Ile Ala Asp Leu Ile Leu
   25                               30                               35                               40

tct gcg ctg gag aga gcc acc gtc ttc cta gaa cag agg ctg cct gaa      198
Ser Ala Leu Glu Arg Ala Thr Val Phe Leu Glu Gln Arg Leu Pro Glu
                               45                               50                               55

atc aac ctg gat ggc atg gtg ggg gtc cga gtg ctg gaa gag cag cta      246
Ile Asn Leu Asp Gly Met Val Gly Val Arg Val Leu Glu Glu Gln Leu
                               60                               65                               70

aaa agt gtc cgg gag aag tgg gcc cag gag ccc ctg ctg caa ccg ctg      294
Lys Ser Val Arg Glu Lys Trp Ala Gln Glu Pro Leu Leu Gln Pro Leu
   75                               80                               85

agc ctg cgc gtg ggg atg ctg ggg gag aag ctg gag gct gcc atc cag      342
Ser Leu Arg Val Gly Met Leu Gly Glu Lys Leu Glu Ala Ala Ile Gln
   90                               95                               100

aga tcc ctc cac tac ctc aag ctg agt gat ccc aag tac cta aga gag      390
Arg Ser Leu His Tyr Leu Lys Leu Ser Asp Pro Lys Tyr Leu Arg Glu
  105                               110                               115                               120

ttc cag ctg acc ctc cag ccc ggg ttt tgg aag ctc cca cat gcc tgg      438
Phe Gln Leu Thr Leu Gln Pro Gly Phe Trp Lys Leu Pro His Ala Trp
                               125                               130                               135

atc cac act gat gcc tcc ttg gtg tac ccc acg ttc ggg ccc cag gac      486
Ile His Thr Asp Ala Ser Leu Val Tyr Pro Thr Phe Gly Pro Gln Asp
                               140                               145                               150

tca ttc tca gag gag aga agt gac gtg tgc ctg gtg cag ctg ctg gga      534
Ser Phe Ser Glu Glu Arg Ser Asp Val Cys Leu Val Gln Leu Leu Gly
   155                               160                               165

acc ggg acg gac agc agc gag ccc tgc ggc ctc tca gac ctc tgc agg      582
Thr Gly Thr Asp Ser Ser Glu Pro Cys Gly Leu Ser Asp Leu Cys Arg
   170                               175                               180

```

agc ctc atg acc aag ccc ggc tgc tca ggc tac tgc ctg tcc cac caa 630
 Ser Leu Met Thr Lys Pro Gly Cys Ser Gly Tyr Cys Leu Ser His Gln
 185 190 195 200

ctg ctc ttc ttc ctc tgg gcc aga atg agg ggg tgc aca cag gga cca 678
 Leu Leu Phe Phe Leu Trp Ala Arg Met Arg Gly Cys Thr Gln Gly Pro
 205 210 215

ctc caa cag agc cag gac tat atc aac ctc ttc tgc gcc aac atg atg 726
 Leu Gln Gln Ser Gln Asp Tyr Ile Asn Leu Phe Cys Ala Asn Met Met
 220 225 230

gac ttg aac cgc aga gct gag gcc atc gga tac gcc tac cct acc cgg 774
 Asp Leu Asn Arg Arg Ala Glu Ala Ile Gly Tyr Ala Tyr Pro Thr Arg
 235 240 245

gac atc ttc atg gaa aac atc atg ttc tgt gga atg ggc ggc ttc tcc 822
 Asp Ile Phe Met Glu Asn Ile Met Phe Cys Gly Met Gly Gly Phe Ser
 250 255 260

gac ttc tac aag ctc cgg tgg ctg gag gcc att ctc agc tgg cag aaa 870
 Asp Phe Tyr Lys Leu Arg Trp Leu Glu Ala Ile Leu Ser Trp Gln Lys
 265 270 275 280

cag cag gaa gga tgc ttc ggg gag cct gat gct gaa gat gaa gaa tta 918
 Gln Gln Glu Gly Cys Phe Gly Glu Pro Asp Ala Glu Asp Glu Glu Leu
 285 290 295

tct aaa gct att caa tat cag cag cat ttt tcg agg aga gtg aag agg 966
 Ser Lys Ala Ile Gln Tyr Gln Gln His Phe Ser Arg Arg Val Lys Arg
 300 305 310

cga gaa aaa caa ttt cca gat ggc tgc tcc tcc cac aac aca gcc aca 1014
 Arg Glu Lys Gln Phe Pro Asp Gly Cys Ser Ser His Asn Thr Ala Thr
 315 320 325

gca gtg gca gcc ctg ggt ggc ttc cta tac atc ctg gca gaa tac ccc 1062
 Ala Val Ala Ala Leu Gly Gly Phe Leu Tyr Ile Leu Ala Glu Tyr Pro
 330 335 340

cca gca aac aga gag cca cac cca tcc aca ccg cca cca cca agc agc 1110
 Pro Ala Asn Arg Glu Pro His Pro Ser Thr Pro Pro Pro Pro Ser Ser
 345 350 355 360

cgc tgagacggac ggttccatgc cagctgcctg gaggaggaac agacccttt 1163
 Arg

agtcctcattc ccttagatcc tggagggcac ggatcacatc ctgggaagaa ggcatctgga 1223
 ggataagcaa agccaccccg acaccaatc ttggaagccc tgagtaggca gggccagggt 1283
 aggtgggggc cgggaggggac ccaggtgtga acggatgaat aaagttcaa 1332

<210> 14

<211> 361

<212> PRT

<213> Homo sapiens

<400> 14

Met Ala Ser Leu Gly Leu Leu Leu Leu Leu Leu Thr Ala Leu Pro
 1 5 10 15
 Pro Leu Trp Ser Ser Ser Leu Pro Gly Leu Asp Thr Ala Glu Ser Lys
 20 25 30
 Ala Thr Ile Ala Asp Leu Ile Leu Ser Ala Leu Glu Arg Ala Thr Val
 35 40 45
 Phe Leu Glu Gln Arg Leu Pro Glu Ile Asn Leu Asp Gly Met Val Gly
 50 55 60
 Val Arg Val Leu Glu Glu Gln Leu Lys Ser Val Arg Glu Lys Trp Ala
 65 70 75 80
 Gln Glu Pro Leu Leu Gln Pro Leu Ser Leu Arg Val Gly Met Leu Gly
 85 90 95
 Glu Lys Leu Glu Ala Ala Ile Gln Arg Ser Leu His Tyr Leu Lys Leu
 100 105 110
 Ser Asp Pro Lys Tyr Leu Arg Glu Phe Gln Leu Thr Leu Gln Pro Gly
 115 120 125
 Phe Trp Lys Leu Pro His Ala Trp Ile His Thr Asp Ala Ser Leu Val
 130 135 140
 Tyr Pro Thr Phe Gly Pro Gln Asp Ser Phe Ser Glu Glu Arg Ser Asp
 145 150 155 160
 Val Cys Leu Val Gln Leu Leu Gly Thr Gly Thr Asp Ser Ser Glu Pro
 165 170 175
 Cys Gly Leu Ser Asp Leu Cys Arg Ser Leu Met Thr Lys Pro Gly Cys
 180 185 190
 Ser Gly Tyr Cys Leu Ser His Gln Leu Leu Phe Phe Leu Trp Ala Arg
 195 200 205
 Met Arg Gly Cys Thr Gln Gly Pro Leu Gln Gln Ser Gln Asp Tyr Ile
 210 215 220
 Asn Leu Phe Cys Ala Asn Met Met Asp Leu Asn Arg Arg Ala Glu Ala
 225 230 235 240
 Ile Gly Tyr Ala Tyr Pro Thr Arg Asp Ile Phe Met Glu Asn Ile Met
 245 250 255
 Phe Cys Gly Met Gly Gly Phe Ser Asp Phe Tyr Lys Leu Arg Trp Leu
 260 265 270
 Glu Ala Ile Leu Ser Trp Gln Lys Gln Gln Glu Gly Cys Phe Gly Glu
 275 280 285
 Pro Asp Ala Glu Asp Glu Glu Leu Ser Lys Ala Ile Gln Tyr Gln Gln
 290 295 300
 His Phe Ser Arg Arg Val Lys Arg Arg Glu Lys Gln Phe Pro Asp Gly
 305 310 315 320
 Cys Ser Ser His Asn Thr Ala Thr Ala Val Ala Ala Leu Gly Gly Phe
 325 330 335
 Leu Tyr Ile Leu Ala Glu Tyr Pro Pro Ala Asn Arg Glu Pro His Pro
 340 345 350
 Ser Thr Pro Pro Pro Pro Ser Ser Arg
 355 360

<210> 15

<211> 1083

<212> DNA

<213> Homo sapiens

<400> 15

atggccagcc	tggggctgct	gctcctgctc	ttactgacag	cactgccacc	gctgtggtcc	60
tcctcactgc	ctgggctgga	cactgctgaa	agtaaagcca	ccattgcaga	cctgatcctg	120
tctgcgctgg	agagagccac	cgtcttccta	gaacagaggc	tgcttgaaat	caacctggat	180
ggcatggtgg	gggtccgagt	gctggaagag	cagctaaaaa	gtgtccggga	gaagtgggcc	240
caggagcccc	tgctgcaacc	gctgagcctg	cgcgtgggga	tgctggggga	gaagctggag	300

```

gctgccatcc agagatccct ccactacctc aagctgagtg atcccaagta cctaagagag 360
ttccagctga ccctccagcc cgggttttgg aagctccac atgcctggat ccacactgat 420
gcctccttgg tgtacccac gttcggggcc caggactcat tctcagagga gagaagtgcac 480
gtgtgcctgg tgcagctgct gggaaccggg acggacagca gcgagccctg cggcctctca 540
gacctctgca ggagcctcat gaccaagccc ggctgctcag gctactgcct gtcccaccaa 600
ctgctcttct tcctctgggc cagaatgagg gggcgcacac agggaccact ccaacagagc 660
caggactata tcaacctctt ctgcgccaac atgatggact tgaaccgcag agctgaggcc 720
atcggaatcg cctaccctac ccgggacatc ttcattgaaa acatcatgtt ctgtggaatg 780
ggcggcttct ccgacttcta caagctccgg tggctggagg ccattctcag ctggcagaaa 840
cagcaggaag gatgcttcgg ggagcctgat gctgaagatg aagaattatc taaagctatt 900
caatatcagc agcatttttc gaggagagtg aagaggcgag aaaaacaatt tccagatggc 960
tgctcctccc acaacacagc cacagcagtg gcagccctgg gtggcttcct atacatcctg 1020
gcagaatacc ccccagcaaa cagagagcca caccatcca caccgccacc accaagcagc 1080
cgc 1083

```

<210> 16

<211> 1400

<212> DNA

<213> Mus musculus

<220>

<221> CDS

<222> (41)...(1195)

<400> 16

```

gtcgaccac gcgtccgcat ccaccagcag aaatcctgtc atg gcg aga ctc ggg 55
                                     Met Ala Arg Leu Gly
                                     1 5

ctg ctt ctc ctc ctg ctg ctg gcc ctg cca cca cac ttc tcc tca gtg 103
Leu Leu Leu Leu Leu Leu Leu Ala Leu Pro Pro His Phe Ser Ser Val
                                     10 15 20

tca tgg cca gac act gca cag ggc acc atg gca aac ttg atc ctg act 151
Ser Trp Pro Asp Thr Ala Gln Gly Thr Met Ala Asn Leu Ile Leu Thr
                                     25 30 35

gca tta gaa aaa gcc acc ttg ttc ttg gag gac agg ctg ccc aca atc 199
Ala Leu Glu Lys Ala Thr Leu Phe Leu Glu Asp Arg Leu Pro Thr Ile
                                     40 45 50

aac ctg gat ggt gtg gtg ggc ttc caa gtg ctg gaa gtg caa ctc cga 247
Asn Leu Asp Gly Val Val Gly Phe Gln Val Leu Glu Val Gln Leu Arg
                                     55 60 65

gga gtt cag gaa aaa tgg gct cac aag ccc ttg ctg cag cct ctc agc 295
Gly Val Gln Glu Lys Trp Ala His Lys Pro Leu Leu Gln Pro Leu Ser
                                     70 75 80 85

atg cgc gct gga cag atg gcc aac aca ctg tct gct ctc ctc caa aaa 343
Met Arg Ala Gly Gln Met Ala Asn Thr Leu Ser Ala Leu Leu Gln Lys
                                     90 95 100

tcc atc ttc tac ctc aag cag agt gac ccc acg tac cta aga gag ttc 391
Ser Ile Phe Tyr Leu Lys Gln Ser Asp Pro Thr Tyr Leu Arg Glu Phe
                                     105 110 115

```

cag cca agc att cag cct ggg ttt tgg aag ttg ccc aat gac tgg aca	439
Gln Pro Ser Ile Gln Pro Gly Phe Trp Lys Leu Pro Asn Asp Trp Thr	
120 125 130	
cgc acc aat gcc tcc cta gtc tac ccc tgg ctg gaa ccc ctg gac tct	487
Arg Thr Asn Ala Ser Leu Val Tyr Pro Trp Leu Glu Pro Leu Asp Ser	
135 140 145	
ttc tca gag gaa agc agc gat gtg tgc ctg gtg caa cta cta gga aca	535
Phe Ser Glu Glu Ser Ser Asp Val Cys Leu Val Gln Leu Leu Gly Thr	
150 155 160 165	
ggg aca gac agc agc cag cct tgc agg ctc tcc aac ttc tgc aga acc	583
Gly Thr Asp Ser Ser Gln Pro Cys Arg Leu Ser Asn Phe Cys Arg Thr	
170 175 180	
ctt atg acc aag gcc ggc tgc tca ggc tac agc ctc tcc cat cag ctg	631
Leu Met Thr Lys Ala Gly Cys Ser Gly Tyr Ser Leu Ser His Gln Leu	
185 190 195	
ctc ttc ttc ctc tgg gcc aga atg caa ggg tgc acg gag gga ctg ttc	679
Leu Phe Phe Leu Trp Ala Arg Met Gln Gly Cys Thr Glu Gly Leu Phe	
200 205 210	
ctc cag agc caa cac tac atg gac atc ttc tgt gcc aat atg atg gaa	727
Leu Gln Ser Gln His Tyr Met Asp Ile Phe Cys Ala Asn Met Met Glu	
215 220 225	
ctg aac cac aga gct gag gcc gtt gga tac gct tac ccc acc caa gac	775
Leu Asn His Arg Ala Glu Ala Val Gly Tyr Ala Tyr Pro Thr Gln Asp	
230 235 240 245	
ctc ttc atg gaa aac att atg ttc tgt ggt atg gct ggc ttc tct gac	823
Leu Phe Met Glu Asn Ile Met Phe Cys Gly Met Ala Gly Phe Ser Asp	
250 255 260	
ttc tac aag ctg cgc tgg ctg gag gcc att ctc agc tgg cag aac ccc	871
Phe Tyr Lys Leu Arg Trp Leu Glu Ala Ile Leu Ser Trp Gln Asn Pro	
265 270 275	
cag gtg gga tgc ttc ggg agg cct gac aca aag ggt gaa cct tct gaa	919
Gln Val Gly Cys Phe Gly Arg Pro Asp Thr Lys Gly Glu Pro Ser Glu	
280 285 290	
gtt cca cat cag cag ggc att ctg aga aga gtg cga agg cgg gaa aaa	967
Val Pro His Gln Gln Gly Ile Leu Arg Arg Val Arg Arg Arg Glu Lys	
295 300 305	
ctg ttc gca gat ggc tgt tgc tgc cac aac aca gcc aca gca gtc gca	1015
Leu Phe Ala Asp Gly Cys Ser Cys His Asn Thr Ala Thr Ala Val Ala	
310 315 320 325	
gcc ctg ggt ggc ttt ctc tac atc ctg gca gaa tac cac cca gac aat	1063
Ala Leu Gly Gly Phe Leu Tyr Ile Leu Ala Glu Tyr His Pro Asp Asn	
330 335 340	

gga gat gca cat cca gaa tac tac cca aac cat gga gat cca tac tca 1111
 Gly Asp Ala His Pro Glu Tyr Tyr Pro Asn His Gly Asp Pro Tyr Ser
 345 350 355

tcc tca cag tca cca gca agc aac tac caa gat ggt gct gcc ggc cct 1159
 Ser Ser Gln Ser Pro Ala Ser Asn Tyr Gln Asp Gly Ala Ala Gly Pro
 360 365 370

gac gtc cag agg act ggc agg ccc ctt agt gtt tct taagtcctga 1205
 Asp Val Gln Arg Thr Gly Arg Pro Leu Ser Val Ser
 375 380 385

gtcagagggtc acaggctgag gaggcaattg aggaaagtga ccagctatat ccccatcgcc 1265
 acttctgggt gtttaaaagt cttgggagag cagggccagg gaaagcaggg ttggagagtg 1325
 ggggtggcca gatgtcagca gaatacataa agcacagtca attggagctg aaaaaaaaaa 1385
 aaaaaggggcg gccgc 1400

<210> 17

<211> 385

<212> PRT

<213> Mus musculus

<400> 17

Met Ala Arg Leu Gly Leu Leu Leu Leu Leu Leu Leu Ala Leu Pro Pro
 1 5 10 15
 His Phe Ser Ser Val Ser Trp Pro Asp Thr Ala Gln Gly Thr Met Ala
 20 25 30
 Asn Leu Ile Leu Thr Ala Leu Glu Lys Ala Thr Leu Phe Leu Glu Asp
 35 40 45
 Arg Leu Pro Thr Ile Asn Leu Asp Gly Val Val Gly Phe Gln Val Leu
 50 55 60
 Glu Val Gln Leu Arg Gly Val Gln Glu Lys Trp Ala His Lys Pro Leu
 65 70 75 80
 Leu Gln Pro Leu Ser Met Arg Ala Gly Gln Met Ala Asn Thr Leu Ser
 85 90 95
 Ala Leu Leu Gln Lys Ser Ile Phe Tyr Leu Lys Gln Ser Asp Pro Thr
 100 105 110
 Tyr Leu Arg Glu Phe Gln Pro Ser Ile Gln Pro Gly Phe Trp Lys Leu
 115 120 125
 Pro Asn Asp Trp Thr Arg Thr Asn Ala Ser Leu Val Tyr Pro Trp Leu
 130 135 140
 Glu Pro Leu Asp Ser Phe Ser Glu Glu Ser Ser Asp Val Cys Leu Val
 145 150 155 160
 Gln Leu Leu Gly Thr Gly Thr Asp Ser Ser Gln Pro Cys Arg Leu Ser
 165 170 175
 Asn Phe Cys Arg Thr Leu Met Thr Lys Ala Gly Cys Ser Gly Tyr Ser
 180 185 190
 Leu Ser His Gln Leu Leu Phe Phe Leu Trp Ala Arg Met Gln Gly Cys
 195 200 205
 Thr Glu Gly Leu Phe Leu Gln Ser Gln His Tyr Met Asp Ile Phe Cys
 210 215 220
 Ala Asn Met Met Glu Leu Asn His Arg Ala Glu Ala Val Gly Tyr Ala
 225 230 235 240
 Tyr Pro Thr Gln Asp Leu Phe Met Glu Asn Ile Met Phe Cys Gly Met
 245 250 255
 Ala Gly Phe Ser Asp Phe Tyr Lys Leu Arg Trp Leu Glu Ala Ile Leu
 260 265 270

Ser Trp Gln Asn Pro Gln Val Gly Cys Phe Gly Arg Pro Asp Thr Lys
 275 280 285
 Gly Glu Pro Ser Glu Val Pro His Gln Gln Gly Ile Leu Arg Arg Val
 290 295 300
 Arg Arg Arg Glu Lys Leu Phe Ala Asp Gly Cys Ser Cys His Asn Thr
 305 310 315 320
 Ala Thr Ala Val Ala Ala Leu Gly Gly Phe Leu Tyr Ile Leu Ala Glu
 325 330 335
 Tyr His Pro Asp Asn Gly Asp Ala His Pro Glu Tyr Tyr Pro Asn His
 340 345 350
 Gly Asp Pro Tyr Ser Ser Ser Gln Ser Pro Ala Ser Asn Tyr Gln Asp
 355 360 365
 Gly Ala Ala Gly Pro Asp Val Gln Arg Thr Gly Arg Pro Leu Ser Val
 370 375 380
 Ser
 385

<210> 18
 <211> 1155
 <212> DNA
 <213> Mus musculus

<400> 18
 atggcgagac tcgggctgct tctcctcctg ctgctggccc tgccaccaca cttctcctca 60
 gtgtcatggc cagacactgc acagggcacc atggcaaact tgatcctgac tgcattagaa 120
 aaagccacct tgttcttgga ggacaggctg cccacaatca acctggatgg tgtggtgggc 180
 ttccaagtgc tggaagtgca actccgagga gttcaggaaa aatgggctca caagcccttg 240
 ctgcagcctc tcagcatgcg cgctggacag atggccaaca cactgtctgc tctcctccaa 300
 aaatccatct tctacctcaa gcagagtgc cccacgtacc taagagagtt ccagccaagc 360
 attcagcctg ggttttggaa gttgcccatt gactggacac gcaccaatgc ctccctagtc 420
 tacccttggc tggaaccctt ggactctttc tcagaggaaa gcagcgatgt gtgcctgggtg 480
 caactactag gaacaggggac agacagcagc cagccttgca ggctctccaa cttctgcaga 540
 acccttatga ccaaggccgg ctgctcaggc tacagcctct cccatcagct gctcttcttc 600
 ctctgggcca gaatgcaagg gtgcacggag ggactgttcc tccagagcca aactacatg 660
 gacatcttct gtgccaatat gatggaactg aaccacagag ctgaggccgt tggatacgt 720
 taccaccacc aagacctctt catggaaaac attatgttct gtggtatggc tggcttctct 780
 gacttctaca agctgcgctg gctggaggcc attctcagct ggcagaacct ccaggtggga 840
 tgcttcggga ggcttgacac aaagggtgaa cttcttgaag ttccacatca gcagggcatt 900
 ctgagaagag tgcgaaggcg ggaaaaactg ttgcagatg gctgttcgtg ccacaacaca 960
 gccacagcag tcgcagccct gggtggcttt ctctacatcc tggcagaata ccaccagac 1020
 aatggagatg cacatccaga atactacca aaccatggag atccatactc atcctcacag 1080
 tcaccagcaa gcaactacca agatggtgct gccggccctg acgtccagag gactggcagg 1140
 ccccttagtg tttct 1155

<210> 19
 <211> 1356
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (105) ... (1148)

<400> 19

gtcgacccac gcgccggca gcctgcagcc cgcagcccgc agcccggagc cagatcgcg	60
gctcagacca aaccgcgactc gaccgcccgc cccagccagg cgcc atg ctg ccg ctt	116
Met Leu Pro Leu	
1	
ctg ctg ggc ctg ctg ggc cca gcg gcc tgc tgg gcc ctg ggc ccg acc	164
Leu Leu Gly Leu Leu Gly Pro Ala Ala Cys Trp Ala Leu Gly Pro Thr	
5 10 15 20	
ccc ggc ccg gga tcc tct gag ctg cgc tcg gcc ttc tcg gcg gca cgc	212
Pro Gly Pro Gly Ser Ser Glu Leu Arg Ser Ala Phe Ser Ala Ala Arg	
25 30 35	
acc acc ccc ctg gag ggc acg tcg gag atg gcg gtg acc ttc gac aag	260
Thr Thr Pro Leu Glu Gly Thr Ser Glu Met Ala Val Thr Phe Asp Lys	
40 45 50	
gtg tac gtg aac atc ggg ggc gac ttc gat gtg gcc acc ggc cag ttt	308
Val Tyr Val Asn Ile Gly Gly Asp Phe Asp Val Ala Thr Gly Gln Phe	
55 60 65	
cgc tgc cgc gtg ccc ggc gcc tac ttc ttc tcc ttc acg gct ggc aag	356
Arg Cys Arg Val Pro Gly Ala Tyr Phe Phe Ser Phe Thr Ala Gly Lys	
70 75 80	
gcc ccg cac aag agc ctg tcg gtg atg ctg gtg cga aac cgc gac gag	404
Ala Pro His Lys Ser Leu Ser Val Met Leu Val Arg Asn Arg Asp Glu	
85 90 95 100	
gtg cag gcg ctg gcc ttc gac gag cag cgg cgg cca ggc gcg cgg cgc	452
Val Gln Ala Leu Ala Phe Asp Glu Gln Arg Arg Pro Gly Ala Arg Arg	
105 110 115	
gca gcc agc cag agc gcc atg ctg cag ctc gac tac ggc gac aca gtg	500
Ala Ala Ser Gln Ser Ala Met Leu Gln Leu Asp Tyr Gly Asp Thr Val	
120 125 130	
tgg ctg cgg ctg cat ggc gcc ccg cag tac gcg cta ggc gcg ccc ggc	548
Trp Leu Arg Leu His Gly Ala Pro Gln Tyr Ala Leu Gly Ala Pro Gly	
135 140 145	
gcc acc ttc agc ggc tac cta gtc tac gcc gac gcc gac gct gac gcg	596
Ala Thr Phe Ser Gly Tyr Leu Val Tyr Ala Asp Ala Asp Ala Asp Ala	
150 155 160	
cct gcg cgc ggg ccg ccc gcg ccc ccc gag ccg cgc tcg gcc ttc tcg	644
Pro Ala Arg Gly Pro Pro Ala Pro Pro Glu Pro Arg Ser Ala Phe Ser	
165 170 175 180	
gcg gcg cgc acg cgc agc ttg gtg ggc tcg gac gct ggc ccc ggg ccg	692
Ala Ala Arg Thr Arg Ser Leu Val Gly Ser Asp Ala Gly Pro Gly Pro	
185 190 195	
cgg cac caa cca ctc gcc ttc gac acc gag ttc gtc aac att ggc ggc	740
Arg His Gln Pro Leu Ala Phe Asp Thr Glu Phe Val Asn Ile Gly Gly	
200 205 210	

gac ttc gac gcg gcg gcc ggc gtg ttc cgc tgc cgt ctg ccc ggc gcc 788
 Asp Phe Asp Ala Ala Ala Gly Val Phe Arg Cys Arg Leu Pro Gly Ala
 215 220 225

tac ttc ttc tcc ttc acg ctg ggc aag ctg ccg cgt aag acg ctg tcg 836
 Tyr Phe Phe Ser Phe Thr Leu Gly Lys Leu Pro Arg Lys Thr Leu Ser
 230 235 240

gtt aag ctg atg aag aac cgc gac gag gtg cag gcc atg att tac gac 884
 Val Lys Leu Met Lys Asn Arg Asp Glu Val Gln Ala Met Ile Tyr Asp
 245 250 255 260

gac ggc gcg tcg cgg cgc cgc gag atg cag agc cag agc gtg atg ctg 932
 Asp Gly Ala Ser Arg Arg Arg Glu Met Gln Ser Gln Ser Val Met Leu
 265 270 275

gcc ctg cgg cgc ggc gac gcc gtc tgg ctg ctc agc cac gac cac gac 980
 Ala Leu Arg Arg Gly Asp Ala Val Trp Leu Leu Ser His Asp His Asp
 280 285 290

ggc tac ggc gcc tac agc aac cac gat ctc cca act gac ctc aaa acg 1028
 Gly Tyr Gly Ala Tyr Ser Asn His Asp Leu Pro Thr Asp Leu Lys Thr
 295 300 305

gtt ttg ccg agt tgg gac gtc cac tgc tgt caa gtc aac cag aga ttt 1076
 Val Leu Pro Ser Trp Asp Val His Cys Cys Gln Val Asn Gln Arg Phe
 310 315 320

gaa ctg tgc att ggt gtg atc cct gag gaa agt cag cac tgg gat gac 1124
 Glu Leu Cys Ile Gly Val Ile Pro Glu Glu Ser Gln His Trp Asp Asp
 325 330 335 340

gcc atc agg atg gat aca gac ctc taactcattg aagcaggaca cctgcacaca 1178
 Ala Ile Arg Met Asp Thr Asp Leu
 345

tgaaagtgag gggagagggg acaaagagct actgaggggaa cagctaactt cagctggagt 1238
 cacctgggtt aatgctgaga gaaaagtcca agcttgggat ggaggaatct gtagtttctt 1298
 tgaaacaagt ctgccactc ccacaaaaaa aaaaaaaaaa aaaaaaaagg gcggccgc 1356

<210> 20

<211> 348

<212> PRT

<213> Homo sapiens

<400> 20

Met Leu Pro Leu Leu Leu Gly Leu Leu Gly Pro Ala Ala Cys Trp Ala
 1 5 10 15
 Leu Gly Pro Thr Pro Gly Pro Gly Ser Ser Glu Leu Arg Ser Ala Phe
 20 25 30
 Ser Ala Ala Arg Thr Thr Pro Leu Gly Thr Ser Glu Met Ala Val
 35 40 45
 Thr Phe Asp Lys Val Tyr Val Asn Ile Gly Gly Asp Phe Asp Val Ala
 50 55 60
 Thr Gly Gln Phe Arg Cys Arg Val Pro Gly Ala Tyr Phe Phe Ser Phe
 65 70 75 80
 Thr Ala Gly Lys Ala Pro His Lys Ser Leu Ser Val Met Leu Val Arg
 85 90 95

Asn Arg Asp Glu Val Gln Ala Leu Ala Phe Asp Glu Gln Arg Arg Pro
 100 105 110
 Gly Ala Arg Arg Ala Ala Ser Gln Ser Ala Met Leu Gln Leu Asp Tyr
 115 120 125
 Gly Asp Thr Val Trp Leu Arg Leu His Gly Ala Pro Gln Tyr Ala Leu
 130 135 140
 Gly Ala Pro Gly Ala Thr Phe Ser Gly Tyr Leu Val Tyr Ala Asp Ala
 145 150 155 160
 Asp Ala Asp Ala Pro Ala Arg Gly Pro Pro Ala Pro Pro Glu Pro Arg
 165 170 175
 Ser Ala Phe Ser Ala Ala Arg Thr Arg Ser Leu Val Gly Ser Asp Ala
 180 185 190
 Gly Pro Gly Pro Arg His Gln Pro Leu Ala Phe Asp Thr Glu Phe Val
 195 200 205
 Asn Ile Gly Gly Asp Phe Asp Ala Ala Ala Gly Val Phe Arg Cys Arg
 210 215 220
 Leu Pro Gly Ala Tyr Phe Phe Ser Phe Thr Leu Gly Lys Leu Pro Arg
 225 230 235 240
 Lys Thr Leu Ser Val Lys Leu Met Lys Asn Arg Asp Glu Val Gln Ala
 245 250 255
 Met Ile Tyr Asp Asp Gly Ala Ser Arg Arg Arg Glu Met Gln Ser Gln
 260 265 270
 Ser Val Met Leu Ala Leu Arg Arg Gly Asp Ala Val Trp Leu Leu Ser
 275 280 285
 His Asp His Asp Gly Tyr Gly Ala Tyr Ser Asn His Asp Leu Pro Thr
 290 295 300
 Asp Leu Lys Thr Val Leu Pro Ser Trp Asp Val His Cys Cys Gln Val
 305 310 315 320
 Asn Gln Arg Phe Glu Leu Cys Ile Gly Val Ile Pro Glu Glu Ser Gln
 325 330 335
 His Trp Asp Asp Ala Ile Arg Met Asp Thr Asp Leu
 340 345

<210> 21

<211> 1044

<212> DNA

<213> Homo sapiens

<400> 21

atgctgccgc	ttctgctggg	cctgctgggc	ccagcggcct	gctgggccc	gggcccgcacc	60
ccgggcccgc	gacccctctga	gctgcgctcg	gccttctcgg	cggaacgcac	cacccccctg	120
gagggcacgt	cggagatggc	ggtgaccttc	gacaagggtgt	acgtgaacat	cgggggacgc	180
ttcgatgtgg	ccaccggcca	gtttcgtcgc	cgcgtgccc	gcgcctactt	cttctccttc	240
acggctggca	aggccccgca	caagagcctg	tcggtgatgc	tggtgcgaaa	ccgcgacgag	300
gtgcaggcgc	tggccttcga	cgagcagcgg	cggccaggcg	cgcggcgcgc	agccagccag	360
agcgccatgc	tgcagctcga	ctacggcgac	acagtgtggc	tgcggctgca	tggcgccccg	420
cagtacgcgc	taggcgcgcc	cggcgccacc	ttcagcggct	acctagtcta	cgccgacgcc	480
gacgctgacg	cgccctgcgc	cgggcgcgcc	gcgcgcgcgc	agccgcgctc	ggccttctcg	540
gcggcgcgca	cgcgcagctt	ggtgggctcg	gacgctggcc	ccgggcgcgc	gcaccaacca	600
ctcgccctcg	acaccgagtt	cgtcaacatt	ggcggcgact	tcgacgcggc	ggccggcggtg	660
ttccgctgcc	gtctgcccgc	cgccacttcc	ttctccttca	cgtgggcaa	gctgccgcgt	720
aagacgctgt	cggttaagct	gatgaagaac	cgcgacgagg	tgcaggccat	gatttacgac	780
gacggcgcg	cgcggcgccg	cgagatgcag	agccagagcg	tgatgctggc	cctgcggcgc	840
ggcgacgcgc	tctggctgct	cagccacgac	cacgacggct	acggcgccca	cagcaaccac	900
gatctcccaa	ctgacctcaa	aacggttttg	ccgagttggg	acgtccactg	ctgtcaagtc	960
aaccagagat	ttgaactgtg	cattgggtgtg	atccctgagg	aaagtcagca	ctgggatgac	1020
gccatcagga	tggatacaga	cctc				1044

<210> 22
 <211> 1388
 <212> DNA
 <213> Catarrhini

<220>
 <221> CDS
 <222> (250)...(1236)

<400> 22
 gtgcacccac ggcgtccgggc tccctccctg gctgaccttc cctttccctc acgcctcccc 60
 acgcccggcc tggggcccca gcacctgtc cgccgcgcc tcagagccgg gagaagcagc 120
 cggagccccc ggcgcccccg ccgcagcgcg ggcggtcagt gcgcagcccg gcgcccgcag 180
 cccgcagccc gcagccggat cgcgggctcg gaccgaaccc gaccggaccg ccgccccag 240
 ccaggctcc atg ctg ccg ctt ctg cta ggc ctg ctg ggc cca gcg gcc tgc 291
 Met Leu Pro Leu Leu Gly Leu Leu Gly Pro Ala Ala Cys
 1 5 10

tgg gcc ctg ggc ccg gcc ccc ggc ccg ggc tcc tca gag ctg cgc tcg 339
 Trp Ala Leu Gly Pro Ala Pro Gly Pro Gly Ser Ser Glu Leu Arg Ser
 15 20 25 30

gcc ttc tcg gcg gca cgc acc act ccc ctg gag ggc gcg tcg gag atg 387
 Ala Phe Ser Ala Ala Arg Thr Thr Pro Leu Glu Gly Ala Ser Glu Met
 35 40 45

gcg gtg acc ttc gac aag gtg tac gtg aac atc ggg ggc gac ttc gac 435
 Ala Val Thr Phe Asp Lys Val Tyr Val Asn Ile Gly Gly Asp Phe Asp
 50 55 60

gcg gcc aca ggc caa ttc cgc tgc cgc gtg ccc ggc gcc tac ttc ttc 483
 Ala Ala Thr Gly Gln Phe Arg Cys Arg Val Pro Gly Ala Tyr Phe Phe
 65 70 75

tcc ttc acg gtt ggc aag gcc ccg cac aag agc ctg tcg gtg atg ctg 531
 Ser Phe Thr Val Gly Lys Ala Pro His Lys Ser Leu Ser Val Met Leu
 80 85 90

gtg cgg aac cac gac gag gtg cag gcg ctg gcc ttc gac gag cag agg 579
 Val Arg Asn His Asp Glu Val Gln Ala Leu Ala Phe Asp Glu Gln Arg
 95 100 105 110

cgg ccc agc gca cgg cgc gcc gcc agc cag agc gcc atg ctg cag ctc 627
 Arg Pro Ser Ala Arg Arg Ala Ala Ser Gln Ser Ala Met Leu Gln Leu
 115 120 125

gac tac ggc gac aca gtt tgg ctg cgg ctg cat ggc gcc ccg cag tac 675
 Asp Tyr Gly Asp Thr Val Trp Leu Arg Leu His Gly Ala Pro Gln Tyr
 130 135 140

gcg ctg ggc gcg ccc ggc gcc acc ttc agc ggc tac ctg gtc tac gcc 723
 Ala Leu Gly Ala Pro Gly Ala Thr Phe Ser Gly Tyr Leu Val Tyr Ala
 145 150 155

gac gcc gac gcc gac gcg cct gcg cgc ggg ccg ccc gcg ccc ccc gag 771
 Asp Ala Asp Ala Asp Ala Pro Ala Arg Gly Pro Pro Ala Pro Pro Glu
 160 165 170

ccg cgc tcg gca ttc tcg gcg gcg cgc acg cgc agc ctg gtg ggc tcg	819
Pro Arg Ser Ala Phe Ser Ala Ala Arg Thr Arg Ser Leu Val Gly Ser	
175 180 185 190	
gac gcg ggg tcc ggg ccg cgg cac cgg ccg cta gcc ttc gac acc gag	867
Asp Ala Gly Ser Gly Pro Arg His Arg Pro Leu Ala Phe Asp Thr Glu	
195 200 205	
ctc gtc aac att ggc ggc gac ttc gac gcg gcg gcc ggc gtg ttc cgc	915
Leu Val Asn Ile Gly Gly Asp Phe Asp Ala Ala Ala Gly Val Phe Arg	
210 215 220	
tgc cgc ctg ccc ggc gcc tac ttc ttc tcc ttc acg ctg ggc aag ctg	963
Cys Arg Leu Pro Gly Ala Tyr Phe Phe Ser Phe Thr Leu Gly Lys Leu	
225 230 235	
ccg cgt aag acg ctg tcg gtg aag ctg atg aag aac cgc gac gag gtg	1011
Pro Arg Lys Thr Leu Ser Val Lys Leu Met Lys Asn Arg Asp Glu Val	
240 245 250	
cag gcc atg att tac gac gac ggc gcg tcg cgg cgc cgc gag atg cag	1059
Gln Ala Met Ile Tyr Asp Asp Gly Ala Ser Arg Arg Arg Glu Met Gln	
255 260 265 270	
agc cag agc gtg atg ctg gcc ctg cgg cgc ggc gac gcc gtc tgg ctg	1107
Ser Gln Ser Val Met Leu Ala Leu Arg Arg Gly Asp Ala Val Trp Leu	
275 280 285	
ctc agc cac gac cac gac ggc tac ggc gcc tac agc aac cac ggc aag	1155
Leu Ser His Asp His Asp Gly Tyr Gly Ala Tyr Ser Asn His Gly Lys	
290 295 300	
tac atc acc ttc tcc ggc ttc ttg gtg tac ccc gac ctc gcc ccc gcc	1203
Tyr Ile Thr Phe Ser Gly Phe Leu Val Tyr Pro Asp Leu Ala Pro Ala	
305 310 315	
gcc ccg cca ggc ctc ggg gcc ccg gaa ctg ctg tgagccccgg gccagagagg	1256
Ala Pro Pro Gly Leu Gly Ala Pro Glu Leu Leu	
320 325	
tgccccgggag ggccaggggc gtgcatgccg ggccaggccc gcggccccgaa cgccctgcgc	1316
gagcgccacg gccgcccggc gcgcctggac tctgccaata aagcagaaag cgggcctgcg	1376
cattgccccg gc	1388
<210> 23	
<211> 329	
<212> PRT	
<213> Catarrhini	
<400> 23	
Met Leu Pro Leu Leu Leu Gly Leu Leu Gly Pro Ala Ala Cys Trp Ala	
1 5 10 15	
Leu Gly Pro Ala Pro Gly Pro Gly Ser Ser Glu Leu Arg Ser Ala Phe	
20 25 30	
Ser Ala Ala Arg Thr Thr Pro Leu Glu Gly Ala Ser Glu Met Ala Val	
35 40 45	
Thr Phe Asp Lys Val Tyr Val Asn Ile Gly Gly Asp Phe Asp Ala Ala	
50 55 60	

Thr Gly Gln Phe Arg Cys Arg Val Pro Gly Ala Tyr Phe Phe Ser Phe
 65 70 75 80
 Thr Val Gly Lys Ala Pro His Lys Ser Leu Ser Val Met Leu Val Arg
 85 90 95
 Asn His Asp Glu Val Gln Ala Leu Ala Phe Asp Glu Gln Arg Arg Pro
 100 105 110
 Ser Ala Arg Arg Ala Ala Ser Gln Ser Ala Met Leu Gln Leu Asp Tyr
 115 120 125
 Gly Asp Thr Val Trp Leu Arg Leu His Gly Ala Pro Gln Tyr Ala Leu
 130 135 140
 Gly Ala Pro Gly Ala Thr Phe Ser Gly Tyr Leu Val Tyr Ala Asp Ala
 145 150 155 160
 Asp Ala Asp Ala Pro Ala Arg Gly Pro Pro Ala Pro Pro Glu Pro Arg
 165 170 175
 Ser Ala Phe Ser Ala Ala Arg Thr Arg Ser Leu Val Gly Ser Asp Ala
 180 185 190
 Gly Ser Gly Pro Arg His Arg Pro Leu Ala Phe Asp Thr Glu Leu Val
 195 200 205
 Asn Ile Gly Gly Asp Phe Asp Ala Ala Ala Gly Val Phe Arg Cys Arg
 210 215 220
 Leu Pro Gly Ala Tyr Phe Phe Ser Phe Thr Leu Gly Lys Leu Pro Arg
 225 230 235 240
 Lys Thr Leu Ser Val Lys Leu Met Lys Asn Arg Asp Glu Val Gln Ala
 245 250 255
 Met Ile Tyr Asp Asp Gly Ala Ser Arg Arg Arg Glu Met Gln Ser Gln
 260 265 270
 Ser Val Met Leu Ala Leu Arg Arg Gly Asp Ala Val Trp Leu Leu Ser
 275 280 285
 His Asp His Asp Gly Tyr Gly Ala Tyr Ser Asn His Gly Lys Tyr Ile
 290 295 300
 Thr Phe Ser Gly Phe Leu Val Tyr Pro Asp Leu Ala Pro Ala Ala Pro
 305 310 315 320
 Pro Gly Leu Gly Ala Pro Glu Leu Leu
 325

<210> 24
 <211> 987
 <212> DNA
 <213> Catarrhini

<400> 24
 atgctgccgc ttctgctagg cctgctgggc ccagcggcct gctgggcccct gggcccggcc 60
 cccggcccgg gctcctcaga gctgcgctcg gccttctcgg cggcacgcac cactcccctg 120
 gagggcgcggt cggagatggc ggtgaccttc gacaaggtgt acgtgaacat cgggggcgac 180
 ttcgacgcgg ccacaggcca attccgctgc cgcgtgcccg gcgcctactt cttctccttc 240
 acggttgcca aggccccgca caagagcctg tcggtgatgc tgggtgcggaa ccacgacgag 300
 gtgcaggcgc tggccttcga cgagcagagg cggcccagcg cacggcgcgcg cgccagccag 360
 agcgccatgc tgcagctcga ctacggcgac acagtttggt tgcggctgca tggcgccccg 420
 cagtaacgcgc tgggcgcgcc cggcgccacc ttcagcgggt acctggtcta cgccgacgcc 480
 gacgccgacg cgctgcgcg cgggcccggc ggcggcccg agccgcgctc ggcattctcg 540
 gcggcgcgca cgcgcagcct ggtgggctcg gacgcggggg ccggggcgcg gcaccggccg 600
 ctacgcttcg acaccgagct cgtcaacatt ggcggcgact tcgacgcggc ggccggcggtg 660
 ttccgctgcc gctgcccgg cgcctacttc ttctccttca cgctgggcaa gctgccgcgt 720
 aagacgctgt cggtgaagct gatgaagaac cgcgacgagg tgcaggccat gatttacgac 780
 gacggcgctg cgggcgccg cgagatgcag agccagagcg tgatgctggc cctgcggcgc 840
 ggcgacgcgg tctgctgct cagccacgac cagcagggct acggcgcccta cagcaaccac 900
 ggcaagtaca tcaccttctc cggcttcttg gtgtaccccg acctcgcccc cgccgccccg 960
 ccaggcctcg gggccccgga actgctg 987

<210> 25
 <211> 625
 <212> DNA
 <213> Mus musculus

```
<400> 25
tcggaccgcc cgccaccagc cacgtgccat gctgctgctc ttgctgggct tcctaggccc      60
ggcggcctgc tgggcactgg gcccggtggt ccctggctcc tcggagctgc ggtcagcctt      120
ctcggcggtc cgcaccaccc cgctggaggg cacgtcggag atggcggtga ccttcgacaa      180
gggttacgtg aacatcgggg gtgacttcga cgcagccacc gggcggttcc gctgtcgcgt      240
gccgggcgcc tactttcttct ccttcacggc cggcaaggcc ccgcacaaaa acctgtcggg      300
gatgctggtg cgcaaccgcg acgaagtgca ggcgctggct ttcgacaagc agcgacggcc      360
aggcgcgcgg cgcgcggcca gccaaagcgc catgctgcag ctcgactacg gcgacacggt      420
gtggctgctg ctgcacggcg ctccgcatta cgcgctcggc gcgccgggcg ccaccttcag      480
cggctacctg gtgtacgcgg acgccgacgc cgacgcgcct gcgcgcgggc ccgcggcccc      540
ggagccgcgc tcggccttct ccgcgcgcca cgccacctgg tgggctccga acccgccccg      600
gcccgcgcca cggcgtttgg ccttc
```

<210> 26
 <211> 26
 <212> PRT
 <213> Homo sapiens

```
<400> 26
Met Ala Glu Leu Pro Gly Pro Phe Leu Cys Gly Ala Leu Leu Gly Phe
  1             5             10             15
Leu Cys Leu Ser Val Pro Pro Ser Asn Pro
             20             25
```

<210> 27
 <211> 136
 <212> PRT
 <213> Homo sapiens

```
<400> 27
Leu Cys Ser Gln Ser Gly Gln Thr Ser Val Gly Gly Ser Thr Ala Leu
  1             5             10             15
Arg Cys Ser Ser Ser Glu Gly Ala Pro Lys Pro Val Tyr Asn Trp Val
             20             25             30
Arg Leu Gly Thr Phe Pro Thr Pro Ser Pro Gly Ser Met Val Gln Asp
             35             40             45
Glu Val Ser Gly Gln Leu Ile Leu Thr Asn Leu Ser Leu Thr Ser Ser
             50             55             60
Gly Thr Tyr Arg Cys Val Ala Thr Asn Gln Met Gly Ser Ala Ser Cys
             65             70             75             80
Glu Leu Thr Leu Ser Val Thr Glu Pro Ser Gln Gly Arg Val Ala Gly
             85             90             95
Ala Leu Ile Gly Val Leu Leu Gly Val Leu Leu Leu Ser Val Ala Ala
             100            105            110
Phe Cys Leu Val Arg Phe Gln Lys Glu Arg Gly Lys Lys Pro Lys Glu
             115            120            125
Thr Tyr Gly Gly Ser Asp Leu Arg
             130            135
```

<210> 28
 <211> 61

<212> PRT

<213> Homo sapiens

<400> 28

```

Gly Gly Ser Thr Ala Leu Arg Cys Ser Ser Ser Glu Gly Ala Pro Lys
 1           5           10           15
Pro Val Tyr Asn Trp Val Arg Leu Gly Thr Phe Pro Thr Pro Ser Pro
          20           25           30
Gly Ser Met Val Gln Asp Glu Val Ser Gly Gln Leu Ile Leu Thr Asn
          35           40           45
Leu Ser Leu Thr Ser Ser Gly Thr Tyr Arg Cys Val Ala
          50           55           60

```

<210> 29

<211> 78

<212> PRT

<213> Artificial sequence

<220>

<223> Synthetically generated peptide

<400> 29

```

Gly Glu Ser Val Thr Leu Thr Cys Ser Val Ser Gly Phe Gly Pro Pro
 1           5           10           15
Gly Val Ser Val Thr Trp Tyr Phe Lys Asn Gly Lys Leu Gly Pro Ser
          20           25           30
Ile Leu Gly Tyr Ser Tyr Ser Arg Ile Glu Ser Gly Glu Lys Ala Asn
          35           40           45
Leu Ser Glu Gly Arg Phe Ser Ile Ser Ser Ile Thr Leu Thr Ile Ser
          50           55           60
Ser Val Glu Lys Glu Asp Ser Gly Thr Tyr Thr Cys Val Val
          65           70           75

```

<210> 30

<211> 189

<212> PRT

<213> Homo sapiens

<400> 30

```

Arg Asn Ser Gln Leu Arg Ile Val Leu Val Gly Lys Thr Gly Ala Gly
 1           5           10           15
Lys Ser Ala Thr Gly Asn Ser Ile Leu Gly Arg Lys Val Phe His Ser
          20           25           30
Gly Thr Ala Ala Lys Ser Ile Thr Lys Lys Cys Glu Lys Arg Ser Ser
          35           40           45
Ser Trp Lys Glu Thr Glu Leu Val Val Val Asp Thr Pro Gly Ile Phe
          50           55           60
Asp Thr Glu Val Pro Asn Ala Glu Thr Ser Lys Glu Ile Ile Arg Cys
          65           70           75           80
Ile Leu Leu Thr Ser Pro Gly Pro His Ala Leu Leu Leu Val Val Pro
          85           90           95
Leu Gly Arg Tyr Thr Glu Glu Glu His Lys Ala Thr Glu Lys Ile Leu
          100          105          110
Lys Met Phe Gly Glu Arg Ala Arg Ser Phe Met Ile Leu Ile Phe Thr
          115          120          125
Arg Lys Asp Asp Leu Gly Asp Thr Asn Leu His Asp Tyr Leu Arg Glu
          130          135          140

```

Ala Pro Glu Asp Ile Gln Asp Leu Met Asp Ile Phe Gly Asp Arg Tyr
 145 150 155 160
 Cys Ala Leu Asn Asn Lys Ala Thr Gly Ala Glu Gln Glu Ala Gln Arg
 165 170 175
 Ala Gln Leu Leu Gly Leu Ile Gln Arg Val Val Arg Glu
 180 185

<210> 31
 <211> 292
 <212> PRT
 <213> Artificial sequence

<220>
 <223> Synthetically generated peptide

<400> 31
 Lys Gly Phe Asp Phe Thr Leu Met Val Val Gly Glu Ser Gly Leu Gly
 1 5 10 15
 Lys Thr Thr Leu Ile Asn Thr Leu Phe Leu Thr Asp Leu Ile Asp Ala
 20 25 30
 Asn Gly Val Ala Asn Asp Ser Arg Glu Ile Asp Gly Ala Ser Glu Thr
 35 40 45
 Lys Ile Lys Lys Thr Val Glu Ile Lys Glu Ile Thr Lys Val Glu Ile
 50 55 60
 Glu Glu Asp Gly Val Lys Leu Asn Leu Thr Val Ile Asp Thr Pro Gly
 65 70 75 80
 Phe Gly Asp Ala Ile Asp Asn Ser Lys Cys Trp Glu Pro Ile Val Glu
 85 90 95
 Tyr Ile Asp Glu Gln His Glu Gln Tyr Leu Arg Gln Glu Ser Arg Ile
 100 105 110
 Asn Arg Thr Lys Ile Val Asp Asn Arg Val His Cys Cys Leu Tyr Phe
 115 120 125
 Ile Ser Pro Thr Gly His Gly Leu Lys Pro Leu Asp Val Glu Phe Met
 130 135 140
 Lys Lys Leu Ser Glu Lys Val Asn Leu Ile Pro Val Ile Ala Lys Ala
 145 150 155 160
 Asp Thr Leu Thr Ala Asp Glu Leu Gln Glu Phe Lys Lys Arg Ile Arg
 165 170 175
 Glu Glu Ile Glu Arg Gln Asn Ile Lys Ile Tyr Lys Phe Pro Asp Glu
 180 185 190
 Glu Glu Asp Glu Gly Asp Glu Glu Phe Lys Glu Gln Thr Gln Gln Leu
 195 200 205
 Lys Ser Ser Ile Pro Phe Ala Ile Val Gly Ser Asn Glu Glu Ile Glu
 210 215 220
 Asn Gly Asp Gly Glu Lys Val Arg Gly Arg Lys Tyr Pro Trp Gly Val
 225 230 235 240
 Val Glu Val Glu Asn Pro Ser His Cys Asp Phe Val Lys Leu Arg Asn
 245 250 255
 Leu Leu Ile Arg Thr His Leu Gln Asp Leu Lys Glu Thr Thr Glu Glu
 260 265 270
 Ile Leu Tyr Glu Asn Tyr Arg Ser Glu Lys Leu Ser Ala Leu Gly Leu
 275 280 285
 Lys Ala Glu Asn
 290

<210> 32
 <211> 163
 <212> PRT

<213> Homo sapiens

<400> 32

Gln Leu Arg Ile Val Leu Val Gly Lys Thr Gly Ala Gly Lys Ser Ala
 1 5 10 15
 Thr Gly Asn Ser Ile Leu Gly Arg Lys Val Phe His Ser Gln Thr Ala
 20 25 30
 Ala Lys Ser Ile Thr Lys Lys Cys Glu Lys Arg Ser Ser Ser Trp Lys
 35 40 45
 Glu Thr Glu Leu Val Val Val Asp Thr Pro Gly Ile Phe Asp Thr Glu
 50 55 60
 Val Pro Asn Ala Glu Thr Ser Lys Glu Ile Ile Arg Cys Ile Leu Leu
 65 70 75 80
 Thr Ser Pro Gly Pro His Ala Leu Leu Leu Val Val Pro Leu Gly Arg
 85 90 95
 Tyr Thr Glu Glu Glu His Lys Ala Thr Glu Lys Ile Leu Lys Met Phe
 100 105 110
 Gly Glu Arg Ala Arg Ser Phe Met Ile Leu Ile Phe Thr Arg Lys Asp
 115 120 125
 Asp Leu Gly Asp Thr Asn Leu His Asp Tyr Leu Arg Glu Ala Pro Glu
 130 135 140
 Asp Ile Gln Asp Leu Met Asp Ile Phe Gly Asp Arg Tyr Cys Ala Leu
 145 150 155 160
 Asn Asn Lys

<210> 33

<211> 198

<212> PRT

<213> Artificial sequence

<220>

<223> Synthetically generated peptide

<400> 33

Gly Glu Val Leu Ala Leu Val Gly Pro Asn Gly Ala Gly Lys Ser Thr
 1 5 10 15
 Leu Leu Lys Leu Ile Ser Gly Leu Leu Pro Pro Thr Glu Gly Thr Ile
 20 25 30
 Leu Leu Asp Gly Ala Arg Asp Leu Arg Leu Ser Lys Leu Lys Glu Arg
 35 40 45
 Leu Glu Arg Leu Arg Lys Asn Ile Gly Val Val Phe Gln Asp Pro Thr
 50 55 60
 Leu Phe Pro Asn Val Glu Leu Thr Val Arg Glu Asn Ile Ala Phe Gly
 65 70 75 80
 Leu Arg Leu Ser Leu Gly Leu Ser Lys Asp Glu Gln Arg Ala Arg Leu
 85 90 95
 Lys Lys Ala Gly Ala Glu Glu Leu Leu Glu Arg Leu Gly Leu Gly Tyr
 100 105 110
 Asp His Leu Leu Asp Arg Arg Pro Gly Thr Leu Ser Gly Gly Gln Lys
 115 120 125
 Gln Arg Val Ala Ile Ala Arg Ala Leu Leu Thr Lys Pro Lys Leu Leu
 130 135 140
 Leu Leu Asp Glu Pro Thr Ala Gly Leu Asp Pro Ala Ser Arg Ala Gln
 145 150 155 160
 Leu Leu Glu Leu Leu Arg Glu Leu Arg Gln Gln Gly Gly Thr Val Leu
 165 170 175

Leu Ile Thr His Asp Leu Asp Leu Leu Asp Arg Leu Ala Asp Arg Ile
 180 185 190
 Leu Val Leu Glu Asp Gly
 195

<210> 34
 <211> 28
 <212> PRT
 <213> Homo sapiens

<400> 34
 Cys Pro Leu Pro Cys Met Asn Gly Gly Gln Cys Ser Ser Arg Asn Gln
 1 5 10 15
 Cys Leu Cys Pro Pro Asp Phe Thr Gly Arg Phe Cys
 20 25

<210> 35
 <211> 36
 <212> PRT
 <213> Homo sapiens

<400> 35
 Cys Ala Met Pro Gly Val Cys Arg His Gly Asp Cys Leu Asn Asn Pro
 1 5 10 15
 Gly Ser Tyr Arg Cys Val Cys Pro Pro Gly His Ser Leu Gly Pro Ser
 20 25 30
 Arg Thr Gln Cys
 35

<210> 36
 <211> 37
 <212> PRT
 <213> Homo sapiens

<400> 36
 Cys Arg Leu Asn Gln Asn Ile Cys Gly His Gly Glu Cys Val Pro Gly
 1 5 10 15
 Pro Pro Asp Tyr Ser Cys His Cys Asn Pro Gly Tyr Arg Ser His Pro
 20 25 30
 Gln His Arg Tyr Cys
 35

<210> 37
 <211> 39
 <212> PRT
 <213> Homo sapiens

<400> 37
 Cys Glu Ala Glu Pro Cys Gly Pro Gly Arg Gly Ile Cys Met Asn Thr
 1 5 10 15
 Gly Gly Ser Tyr Asn Cys His Cys Asn Arg Gly Tyr Arg Leu His Val
 20 25 30
 Gly Ala Gly Gly Arg Ser Cys
 35

<210> 38
 <211> 38
 <212> PRT

<213> Homo sapiens

<400> 38

```

Cys Ala Lys Pro His Leu Cys Gly Asp Gly Gly Phe Cys Ile Asn Phe
 1             5             10             15
Pro Gly His Tyr Lys Cys Asn Cys Tyr Pro Gly Tyr Arg Leu Lys Ala
             20             25             30
Ser Arg Pro Pro Val Cys
             35

```

<210> 39

<211> 36

<212> PRT

<213> Homo sapiens

<400> 39

```

Cys Arg Asp Pro Ser Ser Cys Pro Asp Gly Lys Cys Glu Asn Lys Pro
 1             5             10             15
Gly Ser Phe Lys Cys Ile Ala Cys Gln Pro Gly Tyr Arg Ser Gln Gly
             20             25             30
Gly Gly Ala Cys
             35

```

<210> 40

<211> 36

<212> PRT

<213> Homo sapiens

<400> 40

```

Cys Ala Glu Gly Ser Pro Cys Ser Pro Gly Trp Cys Glu Asn Leu Pro
 1             5             10             15
Gly Ser Phe Arg Cys Thr Cys Ala Gln Gly Tyr Ala Pro Ala Pro Asp
             20             25             30
Gly Arg Ser Cys
             35

```

<210> 41

<211> 36

<212> PRT

<213> Homo sapiens

<400> 41

```

Cys Glu Ala Gly Asp Val Cys Asp Asn Gly Ile Cys Ser Asn Thr Pro
 1             5             10             15
Gly Ser Phe Gln Cys Gln Cys Leu Ser Gly Tyr His Leu Ser Arg Asp
             20             25             30
Arg Ser His Cys
             35

```

<210> 42

<211> 35

<212> PRT

<213> Homo sapiens

<400> 42

```

Cys Asp Phe Pro Ala Ala Cys Ile Gly Gly Asp Cys Ile Asn Thr Asn
 1             5             10             15

```

Gly Ser Tyr Arg Cys Leu Cys Pro Gln Gly His Arg Leu Val Gly Gly
 20 25 30
 Arg Lys Cys
 35

<210> 43
 <211> 38
 <212> PRT
 <213> Homo sapiens

<400> 43
 Cys Ser Gln Asp Pro Ser Leu Cys Leu Pro His Gly Ala Cys Lys Asn
 1 5 10 15
 Leu Gln Gly Ser Tyr Val Cys Val Cys Asp Glu Gly Phe Thr Pro Thr
 20 25 30
 Gln Asp Gln His Gly Cys
 35

<210> 44
 <211> 38
 <212> PRT
 <213> Homo sapiens

<400> 44
 Cys Met Leu Phe Gly Ser Glu Ile Cys Lys Glu Gly Lys Cys Val Asn
 1 5 10 15
 Thr Gln Pro Gly Tyr Glu Cys Tyr Cys Lys Gln Gly Phe Tyr Tyr Asp
 20 25 30
 Gly Asn Leu Leu Glu Cys
 35

<210> 45
 <211> 35
 <212> PRT
 <213> Homo sapiens

<400> 45
 Cys Leu Asp Glu Ser Asn Cys Arg Asn Gly Val Cys Glu Asn Thr Arg
 1 5 10 15
 Gly Gly Tyr Arg Cys Ala Cys Thr Pro Pro Ala Glu Tyr Ser Pro Ala
 20 25 30
 Gln Arg Gln
 35

<210> 46
 <211> 36
 <212> PRT
 <213> Homo sapiens

<400> 46
 Cys Gln Asp Pro Ala Ala Cys Arg Pro Gly Arg Cys Val Asn Leu Pro
 1 5 10 15
 Gly Ser Tyr Arg Cys Glu Cys Arg Pro Pro Trp Val Pro Gly Pro Ser
 20 25 30
 Gly Arg Asp Cys
 35

<210> 47

<211> 36
 <212> PRT
 <213> Homo sapiens

<400> 47
 Asp Ser Asp Glu Cys Arg Cys Val Ser Gly Arg Cys Val Pro Arg Pro
 1 5 10 15
 Gly Gly Ala Ala Cys Glu Cys Pro Gly Gly Phe Gln Leu Asp Ala Ser
 20 25 30
 Arg Ala Arg Cys
 35

<210> 48
 <211> 40
 <212> PRT
 <213> Homo sapiens

<400> 48
 Cys Arg Glu Leu Asn Gln Arg Gly Ile Leu Cys Lys Ser Glu Arg Cys
 1 5 10 15
 Val Asn Thr Ser Gly Ser Phe Arg Cys Val Cys Lys Ala Gly Phe Ala
 20 25 30
 Arg Ser Arg Pro His Gly Ala Cys
 35 40

<210> 49
 <211> 43
 <212> PRT
 <213> Artificial sequence

<220>
 <223> Synthetically generated peptide

<400> 49
 Cys Asn Pro Asn Thr Gly Pro Cys Leu Asn Gly Gly Thr Cys Val Asn
 1 5 10 15
 Thr Pro Gly Gly Ser Val Phe Gly Gly Tyr Thr Cys Glu Cys Pro Glu
 20 25 30
 Gly Tyr Ala Leu Ser Tyr Thr Gly Lys Arg Cys
 35 40

<210> 50
 <211> 44
 <212> PRT
 <213> Homo sapiens

<400> 50
 Gln Pro Cys Gly Ser Asn Pro Leu Pro Gly Leu Thr Lys Gln Glu Asp
 1 5 10 15
 Cys Cys Gly Ser Ile Gly Thr Ala Trp Gly Gln Ser Lys Cys His Lys
 20 25 30
 Cys Pro Gln Leu Gln Tyr Thr Gly Val Gln Lys Pro
 35 40

<210> 51
 <211> 42
 <212> PRT
 <213> Homo sapiens

<400> 51
 His Gln Cys Gln His Pro Leu Thr Thr Arg Leu Thr Arg Gln Leu Cys
 1 5 10 15
 Cys Cys Ser Val Gly Lys Ala Trp Gly Ala Arg Cys Gln Arg Cys Pro
 20 25 30
 Thr Asp Gly Thr Ala Ala Phe Lys Glu Ile
 35 40

<210> 52
 <211> 44
 <212> PRT
 <213> Homo sapiens

<400> 52
 Val Phe Cys Asp Ser Val Leu Ala Thr Asn Val Thr Gln Gln Glu Cys
 1 5 10 15
 Cys Cys Ser Leu Gly Ala Gly Trp Gly Asp His Cys Glu Ile Tyr Pro
 20 25 30
 Cys Pro Val Tyr Ser Ser Ala Glu Phe His Ser Leu
 35 40

<210> 53
 <211> 46
 <212> PRT
 <213> Homo sapiens

<400> 53
 Gly Met Cys Ala Gly Pro Leu Ala Gly Pro Ala Leu Thr Phe Asp Asp
 1 5 10 15
 Cys Cys Cys Arg Gln Gly Arg Gly Trp Gly Ala Gln Cys Arg Pro Cys
 20 25 30
 Pro Pro Arg Gly Ala Gly Ser His Cys Pro Thr Ser Gln Ser
 35 40 45

<210> 54
 <211> 45
 <212> PRT
 <213> Artificial sequence

<220>
 <223> Synthetically generated peptide

<400> 54
 Gly Arg Cys Ser Asn Pro Leu Pro Gly Arg Ala Val Thr Lys Ser Glu
 1 5 10 15
 Cys Cys Cys Ser Val Gly Arg Gly Glu Ala Trp Gly Thr Pro Cys Glu
 20 25 30
 Leu Cys Pro Val Pro Gly Thr Ala Glu Phe Lys Glu Leu
 35 40 45

<210> 55
 <211> 65
 <212> PRT
 <213> Homo sapiens

<400> 55
 Arg Asp Pro Ser Ser Cys Pro Asp Gly Lys Cys Glu Asn Lys Pro Gly
 1 5 10 15

Ser Phe Lys Cys Ile Ala Cys Gln Pro Gly Tyr Arg Ser Gln Gly Gly
 20 25 30
 Gly Ala Cys Arg Asp Val Asn Glu Cys Ala Glu Gly Ser Pro Cys Ser
 35 40 45
 Pro Gly Trp Cys Glu Asn Leu Pro Gly Ser Phe Arg Cys Thr Cys Ala
 50 55 60
 Gln
 65

<210> 56

<211> 67

<212> PRT

<213> Artificial sequence

<220>

<223> Synthetically generated peptide

<400> 56

Met Asp Pro Gln Asn Cys Ser Cys Ala Thr Gly Gly Ser Cys Thr Cys
 1 5 10 15
 Gly Thr Ser Cys Lys Cys Lys Asn Cys Lys Cys Thr Ser Cys Lys Lys
 20 25 30
 Ser Cys Cys Ser Cys Cys Pro Ala Gly Cys Ser Lys Cys Ala Gly Gly
 35 40 45
 Cys Val Cys Lys Gly Gly Gly Ala Ala Ser Glu Thr Ser Lys Cys Ser
 50 55 60
 Cys Cys Ala
 65

<210> 57

<211> 305

<212> PRT

<213> Homo sapiens

<400> 57

Met Ala Ala Gln Tyr Gly Ser Met Ser Phe Asn Pro Ser Thr Pro Gly
 1 5 10 15
 Ala Ser Tyr Gly Pro Gly Arg Gln Glu Pro Arg Asn Ser Gln Leu Arg
 20 25 30
 Ile Val Leu Val Gly Lys Thr Gly Ala Gly Lys Ser Ala Thr Gly Asn
 35 40 45
 Ser Ile Leu Gly Arg Lys Val Phe His Ser Gly Thr Ala Ala Lys Ser
 50 55 60
 Ile Thr Lys Lys Cys Glu Lys Arg Ser Ser Ser Trp Lys Glu Thr Glu
 65 70 75 80
 Leu Val Val Val Asp Thr Pro Gly Ile Phe Asp Thr Glu Val Pro Asn
 85 90 95
 Ala Glu Thr Ser Lys Glu Ile Ile Arg Cys Ile Leu Leu Thr Ser Pro
 100 105 110
 Gly Pro His Ala Leu Leu Leu Val Pro Leu Gly Arg Tyr Thr Glu
 115 120 125
 Glu Glu His Lys Ala Thr Glu Lys Ile Leu Lys Met Phe Gly Glu Arg
 130 135 140
 Ala Arg Ser Phe Met Ile Leu Ile Phe Thr Arg Lys Asp Asp Leu Gly
 145 150 155 160
 Asp Thr Asn Leu His Asp Tyr Leu Arg Glu Ala Pro Glu Asp Ile Gln
 165 170 175

Asp Leu Met Asp Ile Phe Gly Asp Arg Tyr Cys Ala Leu Asn Asn Lys
 180 185 190
 Ala Thr Gly Ala Glu Gln Glu Ala Gln Arg Ala Gln Leu Leu Gly Leu
 195 200 205
 Ile Gln Arg Val Val Arg Glu Asn Lys Glu Gly Cys Tyr Thr Asn Arg
 210 215 220
 Met Tyr Gln Arg Ala Glu Glu Glu Ile Gln Lys Gln Thr Gln Ala Met
 225 230 235 240
 Gln Glu Leu His Arg Val Glu Leu Glu Arg Glu Lys Ala Arg Ile Arg
 245 250 255
 Glu Glu Tyr Glu Glu Lys Ile Arg Lys Leu Glu Asp Lys Val Glu Gln
 260 265 270
 Glu Lys Arg Lys Lys Gln Met Glu Lys Lys Leu Ala Glu Gln Glu Ala
 275 280 285
 His Tyr Ala Val Arg Gln Gln Arg Ala Arg Thr Glu Val Glu Ser Lys
 290 295 300
 Asp
 305

<210> 58
 <211> 18
 <212> PRT
 <213> Homo sapiens

<400> 58
 Gly Ile Leu Glu Leu Ile Met Thr Ala Leu Gln Ile Ala Ser Phe Ile
 1 5 10 15
 Leu Leu

<210> 59
 <211> 6
 <212> PRT
 <213> Homo sapiens

<400> 59
 Arg Leu Phe Ala Glu Asp
 1 5

<210> 60
 <211> 29
 <212> PRT
 <213> Homo sapiens

<400> 60
 Met Arg Gly Ala Gly Ala Ala Gly Leu Leu Ala Leu Leu Leu Leu
 1 5 10 15
 Leu Leu Leu Leu Leu Gly Leu Gly Gly Arg Val Glu Gly
 20 25

<210> 61
 <211> 1260
 <212> PRT
 <213> Homo sapiens

<400> 61
 Gly Pro Ala Gly Glu Arg Gly Ala Gly Gly Gly Gly Ala Leu Ala Arg
 1 5 10 15

Glu Arg Phe Lys Val Val Phe Ala Pro Val Ile Cys Lys Arg Thr Cys
 20 25 30
 Leu Lys Gly Gln Cys Arg Asp Ser Cys Gln Gln Gly Ser Asn Met Thr
 35 40 45
 Leu Ile Gly Glu Asn Gly His Ser Thr Asp Thr Leu Thr Gly Ser Gly
 50 55 60
 Phe Arg Val Val Val Cys Pro Leu Pro Cys Met Asn Gly Gly Gln Cys
 65 70 75 80
 Ser Ser Arg Asn Gln Cys Leu Cys Pro Pro Asp Phe Thr Gly Arg Phe
 85 90 95
 Cys Gln Val Pro Gln Ala Gly Gly Ala Gly Gly Gly Thr Gly Gly Ser
 100 105 110
 Gly Pro Gly Leu Ser Arg Thr Gly Ala Leu Ser Thr Gly Ala Leu Pro
 115 120 125
 Pro Leu Ala Pro Glu Gly Asp Ser Val Ala Ser Lys His Ala Ile Tyr
 130 135 140
 Ala Val Gln Val Ile Ala Asp Pro Pro Gly Pro Gly Glu Gly Pro Pro
 145 150 155 160
 Ala Gln His Ala Ala Phe Leu Val Pro Leu Gly Pro Gly Gln Ile Ser
 165 170 175
 Ala Glu Val Gln Ala Pro Pro Pro Val Val Asn Val Arg Val His His
 180 185 190
 Pro Pro Glu Ala Ser Val Gln Val His Arg Ile Glu Ser Ser Asn Ala
 195 200 205
 Glu Ser Ala Ala Pro Ser Gln His Leu Leu Pro His Pro Lys Pro Ser
 210 215 220
 His Pro Arg Pro Pro Thr Gln Lys Pro Leu Gly Arg Cys Phe Gln Asp
 225 230 235 240
 Thr Leu Pro Lys Gln Pro Cys Gly Ser Asn Pro Leu Pro Gly Leu Thr
 245 250 255
 Lys Gln Glu Asp Cys Cys Gly Ser Ile Gly Thr Ala Trp Gly Gln Ser
 260 265 270
 Lys Cys His Lys Cys Pro Gln Leu Gln Tyr Thr Gly Val Gln Lys Pro
 275 280 285
 Gly Pro Val Arg Gly Glu Val Gly Ala Asp Cys Pro Gln Gly Tyr Lys
 290 295 300
 Arg Leu Asn Ser Thr His Cys Gln Asp Ile Asn Glu Cys Ala Met Pro
 305 310 315 320
 Gly Val Cys Arg His Gly Asp Cys Leu Asn Asn Pro Gly Ser Tyr Arg
 325 330 335
 Cys Val Cys Pro Pro Gly His Ser Leu Gly Pro Ser Arg Thr Gln Cys
 340 345 350
 Ile Ala Asp Lys Pro Glu Glu Lys Ser Leu Cys Phe Arg Leu Val Ser
 355 360 365
 Pro Glu His Gln Cys Gln His Pro Leu Thr Thr Arg Leu Thr Arg Gln
 370 375 380
 Leu Cys Cys Cys Ser Val Gly Lys Ala Trp Gly Ala Arg Cys Gln Arg
 385 390 395 400
 Cys Pro Thr Asp Gly Thr Ala Ala Phe Lys Glu Ile Cys Pro Ala Gly
 405 410 415
 Lys Gly Tyr His Ile Leu Thr Ser His Gln Thr Leu Thr Ile Gln Gly
 420 425 430
 Glu Ser Asp Phe Ser Leu Phe Leu His Pro Asp Gly Pro Pro Lys Pro
 435 440 445
 Gln Gln Leu Pro Glu Ser Pro Ser Gln Ala Pro Pro Pro Glu Asp Thr
 450 455 460
 Glu Glu Arg Gly Val Thr Thr Asp Ser Pro Val Ser Glu Glu Arg Ser
 465 470 475 480

Val Gln Gln Ser His Pro Thr Ala Thr Thr Thr Pro Ala Arg Pro Tyr
 485 490 495
 Pro Glu Leu Ile Ser Arg Pro Ser Pro Pro Thr Met Arg Trp Phe Leu
 500 505 510
 Pro Asp Leu Pro Pro Ser Arg Ser Ala Val Glu Ile Ala Pro Thr Gln
 515 520 525
 Val Thr Glu Thr Asp Glu Cys Arg Leu Asn Gln Asn Ile Cys Gly His
 530 535 540
 Gly Glu Cys Val Pro Gly Pro Pro Asp Tyr Ser Cys His Cys Asn Pro
 545 550 555 560
 Gly Tyr Arg Ser His Pro Gln His Arg Tyr Cys Val Asp Val Asn Glu
 565 570 575
 Cys Glu Ala Glu Pro Cys Gly Pro Gly Arg Gly Ile Cys Met Asn Thr
 580 585 590
 Gly Gly Ser Tyr Asn Cys His Cys Asn Arg Gly Tyr Arg Leu His Val
 595 600 605
 Gly Ala Gly Gly Arg Ser Cys Val Asp Leu Asn Glu Cys Ala Lys Pro
 610 615 620
 His Leu Cys Gly Asp Gly Gly Phe Cys Ile Asn Phe Pro Gly His Tyr
 625 630 635 640
 Lys Cys Asn Cys Tyr Pro Gly Tyr Arg Leu Lys Ala Ser Arg Pro Pro
 645 650 655
 Val Cys Glu Asp Ile Asp Glu Cys Arg Asp Pro Ser Ser Cys Pro Asp
 660 665 670
 Gly Lys Cys Glu Asn Lys Pro Gly Ser Phe Lys Cys Ile Ala Cys Gln
 675 680 685
 Pro Gly Tyr Arg Ser Gln Gly Gly Gly Ala Cys Arg Asp Val Asn Glu
 690 695 700
 Cys Ala Glu Gly Ser Pro Cys Ser Pro Gly Trp Cys Glu Asn Leu Pro
 705 710 715 720
 Gly Ser Phe Arg Cys Thr Cys Ala Gln Gly Tyr Ala Pro Ala Pro Asp
 725 730 735
 Gly Arg Ser Cys Leu Asp Val Asp Glu Cys Glu Ala Gly Asp Val Cys
 740 745 750
 Asp Asn Gly Ile Cys Ser Asn Thr Pro Gly Ser Phe Gln Cys Gln Cys
 755 760 765
 Leu Ser Gly Tyr His Leu Ser Arg Asp Arg Ser His Cys Glu Asp Ile
 770 775 780
 Asp Glu Cys Asp Phe Pro Ala Ala Cys Ile Gly Gly Asp Cys Ile Asn
 785 790 795 800
 Thr Asn Gly Ser Tyr Arg Cys Leu Cys Pro Gln Gly His Arg Leu Val
 805 810 815
 Gly Gly Arg Lys Cys Gln Asp Ile Asp Glu Cys Ser Gln Asp Pro Ser
 820 825 830
 Leu Cys Leu Pro His Gly Ala Cys Lys Asn Leu Gln Gly Ser Tyr Val
 835 840 845
 Cys Val Cys Asp Glu Gly Phe Thr Pro Thr Gln Asp Gln His Gly Cys
 850 855 860
 Glu Glu Val Glu Gln Pro His His Lys Lys Glu Cys Tyr Leu Asn Phe
 865 870 875 880
 Asp Asp Thr Val Phe Cys Asp Ser Val Leu Ala Thr Asn Val Thr Gln
 885 890 895
 Gln Glu Cys Cys Cys Ser Leu Gly Ala Gly Trp Gly Asp His Cys Glu
 900 905 910
 Ile Tyr Pro Cys Pro Val Tyr Ser Ser Ala Glu Phe His Ser Leu Cys
 915 920 925
 Pro Asp Gly Lys Gly Tyr Thr Gln Asp Asn Asn Ile Val Asn Tyr Gly
 930 935 940

```

Ile Pro Ala His Arg Asp Ile Asp Glu Cys Met Leu Phe Gly Ser Glu
945                      950                      955                      960
Ile Cys Lys Glu Gly Lys Cys Val Asn Thr Gln Pro Gly Tyr Glu Cys
                      965                      970                      975
Tyr Cys Lys Gln Gly Phe Tyr Tyr Asp Gly Asn Leu Leu Glu Cys Val
                      980                      985                      990
Asp Val Asp Glu Cys Leu Asp Glu Ser Asn Cys Arg Asn Gly Val Cys
                      995                      1000                      1005
Glu Asn Thr Arg Gly Gly Tyr Arg Cys Ala Cys Thr Pro Pro Ala Glu
1010                      1015                      1020
Tyr Ser Pro Ala Gln Arg Gln Cys Leu Ser Pro Glu Glu Met Asp Val
1025                      1030                      1035                      1040
Asp Glu Cys Gln Asp Pro Ala Ala Cys Arg Pro Gly Arg Cys Val Asn
                      1045                      1050                      1055
Leu Pro Gly Ser Tyr Arg Cys Glu Cys Arg Pro Pro Trp Val Pro Gly
1060                      1065                      1070
Pro Ser Gly Arg Asp Cys Gln Leu Pro Glu Ser Pro Ala Glu Arg Ala
1075                      1080                      1085
Pro Glu Arg Arg Asp Val Cys Trp Ser Gln Arg Gly Glu Asp Gly Met
1090                      1095                      1100
Cys Ala Gly Pro Leu Ala Gly Pro Ala Leu Thr Phe Asp Asp Cys Cys
1105                      1110                      1115                      1120
Cys Arg Gln Gly Arg Gly Trp Gly Ala Gln Cys Arg Pro Cys Pro Pro
                      1125                      1130                      1135
Arg Gly Ala Gly Ser His Cys Pro Thr Ser Gln Ser Glu Ser Asn Ser
1140                      1145                      1150
Phe Trp Asp Thr Ser Pro Leu Leu Leu Gly Lys Pro Pro Arg Asp Glu
1155                      1160                      1165
Asp Ser Ser Glu Glu Asp Ser Asp Glu Cys Arg Cys Val Ser Gly Arg
1170                      1175                      1180
Cys Val Pro Arg Pro Gly Gly Ala Ala Cys Glu Cys Pro Gly Gly Phe
1185                      1190                      1195                      1200
Gln Leu Asp Ala Ser Arg Ala Arg Cys Val Asp Ile Asp Glu Cys Arg
                      1205                      1210                      1215
Glu Leu Asn Gln Arg Gly Leu Leu Cys Lys Ser Glu Arg Cys Val Asn
1220                      1225                      1230
Thr Ser Gly Ser Phe Arg Cys Val Cys Lys Ala Gly Phe Ala Arg Ser
1235                      1240                      1245
Arg Pro His Gly Ala Cys Val Pro Gln Arg Arg Arg
1250                      1255                      1260

```

<210> 62
 <211> 20
 <212> PRT
 <213> Homo sapiens

<400> 62
 Met Ala Ser Leu Gly Leu Leu Leu Leu Leu Leu Thr Ala Leu Pro
 1 5 10 15
 Pro Leu Trp Ser
 20

<210> 63
 <211> 341
 <212> PRT
 <213> Homo sapiens

<400> 63

Ser Ser Leu Pro Gly Leu Asp Thr Ala Glu Ser Lys Ala Thr Ile Ala
 1 5 10 15
 Asp Leu Ile Leu Ser Ala Leu Glu Arg Ala Thr Val Phe Leu Glu Gln
 20 25 30
 Arg Leu Pro Glu Ile Asn Leu Asp Gly Met Val Gly Val Arg Val Leu
 35 40 45
 Glu Glu Gln Leu Lys Ser Val Arg Glu Lys Trp Ala Gln Glu Pro Leu
 50 55 60
 Leu Gln Pro Leu Ser Leu Arg Val Gly Met Leu Gly Glu Lys Leu Glu
 65 70 75 80
 Ala Ala Ile Gln Arg Ser Leu His Tyr Leu Lys Leu Ser Asp Pro Lys
 85 90 95
 Tyr Leu Arg Glu Phe Gln Leu Thr Leu Gln Pro Gly Phe Trp Lys Leu
 100 105 110
 Pro His Ala Trp Ile His Thr Asp Ala Ser Leu Val Tyr Pro Thr Phe
 115 120 125
 Gly Pro Gln Asp Ser Phe Ser Glu Glu Arg Ser Asp Val Cys Leu Val
 130 135 140
 Gln Leu Leu Gly Thr Gly Thr Asp Ser Ser Glu Pro Cys Gly Leu Ser
 145 150 155 160
 Asp Leu Cys Arg Ser Leu Met Thr Lys Pro Gly Cys Ser Gly Tyr Cys
 165 170 175
 Leu Ser His Gln Leu Leu Phe Phe Leu Trp Ala Arg Met Arg Gly Cys
 180 185 190
 Thr Gln Gly Pro Leu Gln Gln Ser Gln Asp Tyr Ile Asn Leu Phe Cys
 195 200 205
 Ala Asn Met Met Asp Leu Asn Arg Arg Ala Glu Ala Ile Gly Tyr Ala
 210 215 220
 Tyr Pro Thr Arg Asp Ile Phe Met Glu Asn Ile Met Phe Cys Gly Met
 225 230 235 240
 Gly Gly Phe Ser Asp Phe Tyr Lys Leu Arg Trp Leu Glu Ala Ile Leu
 245 250 255
 Ser Trp Gln Lys Gln Gln Glu Gly Cys Phe Gly Glu Pro Asp Ala Glu
 260 265 270
 Asp Glu Glu Leu Ser Lys Ala Ile Gln Tyr Gln Gln His Phe Ser Arg
 275 280 285
 Arg Val Lys Arg Arg Glu Lys Gln Phe Pro Asp Gly Cys Ser Ser His
 290 295 300
 Asn Thr Ala Thr Ala Val Ala Ala Leu Gly Gly Phe Leu Tyr Ile Leu
 305 310 315 320
 Ala Glu Tyr Pro Pro Ala Asn Arg Glu Pro His Pro Ser Thr Pro Pro
 325 330 335
 Pro Pro Ser Ser Arg
 340

<210> 64

<211> 19

<212> PRT

<213> Mus musculus

<400> 64

Met Ala Arg Leu Gly Leu Leu Leu Leu Leu Leu Ala Leu Pro Pro
 1 5 10 15
 His Phe Ser

<210> 65

<211> 366

<212> PRT

<213> Mus musculus

<400> 65

```

Ser Val Ser Trp Pro Asp Thr Ala Gln Gly Thr Met Ala Asn Leu Ile
 1           5           10           15
Leu Thr Ala Leu Glu Lys Ala Thr Leu Phe Leu Glu Asp Arg Leu Pro
      20           25           30
Thr Ile Asn Leu Asp Gly Val Val Gly Phe Gln Val Leu Glu Val Gln
      35           40           45
Leu Arg Gly Val Gln Glu Lys Trp Ala His Lys Pro Leu Leu Gln Pro
      50           55           60
Leu Ser Met Arg Ala Gly Gln Met Ala Asn Thr Leu Ser Ala Leu Leu
      65           70           75           80
Gln Lys Ser Ile Phe Tyr Leu Lys Gln Ser Asp Pro Thr Tyr Leu Arg
      85           90           95
Glu Phe Gln Pro Ser Ile Gln Pro Gly Phe Trp Lys Leu Pro Asn Asp
      100          105          110
Trp Thr Arg Thr Asn Ala Ser Leu Val Tyr Pro Trp Leu Glu Pro Leu
      115          120          125
Asp Ser Phe Ser Glu Glu Ser Ser Asp Val Cys Leu Val Gln Leu Leu
      130          135          140
Gly Thr Gly Thr Asp Ser Ser Gln Pro Cys Arg Leu Ser Asn Phe Cys
      145          150          155          160
Arg Thr Leu Met Thr Lys Ala Gly Cys Ser Gly Tyr Ser Leu Ser His
      165          170          175
Gln Leu Leu Phe Phe Leu Trp Ala Arg Met Gln Gly Cys Thr Glu Gly
      180          185          190
Leu Phe Leu Gln Ser Gln His Tyr Met Asp Ile Phe Cys Ala Asn Met
      195          200          205
Met Glu Leu Asn His Arg Ala Glu Ala Val Gly Tyr Ala Tyr Pro Thr
      210          215          220
Gln Asp Leu Phe Met Glu Asn Ile Met Phe Cys Gly Met Ala Gly Phe
      225          230          235          240
Ser Asp Phe Tyr Lys Leu Arg Trp Leu Glu Ala Ile Leu Ser Trp Gln
      245          250          255
Asn Pro Gln Val Gly Cys Phe Gly Arg Pro Asp Thr Lys Gly Glu Pro
      260          265          270
Ser Glu Val Pro His Gln Gln Gly Ile Leu Arg Arg Val Arg Arg Arg
      275          280          285
Glu Lys Leu Phe Ala Asp Gly Cys Ser Cys His Asn Thr Ala Thr Ala
      290          295          300
Val Ala Ala Leu Gly Gly Phe Leu Tyr Ile Leu Ala Glu Tyr His Pro
      305          310          315          320
Asp Asn Gly Asp Ala His Pro Glu Tyr Tyr Pro Asn His Gly Asp Pro
      325          330          335
Tyr Ser Ser Ser Gln Ser Pro Ala Ser Asn Tyr Gln Asp Gly Ala Ala
      340          345          350
Gly Pro Asp Val Gln Arg Thr Gly Arg Pro Leu Ser Val Ser
      355          360          365

```

<210> 66

<211> 14

<212> PRT

<213> Homo sapiens

<400> 66

Met Leu Pro Leu Leu Leu Gly Leu Leu Gly Pro Ala Ala Cys
 1 5 10

<210> 67
 <211> 334
 <212> PRT
 <213> Homo sapiens

<400> 67
 Trp Ala Leu Gly Pro Thr Pro Gly Pro Gly Ser Ser Glu Leu Arg Gly
 1 5 10 15
 Ala Phe Ser Ala Ala Arg Thr Thr Pro Leu Glu Gly Thr Ser Glu Met
 20 25 30
 Ala Val Thr Phe Asp Lys Val Tyr Val Asn Ile Gly Gly Asp Phe Asp
 35 40 45
 Val Ala Thr Gly Gln Phe Arg Cys Arg Val Pro Gly Ala Tyr Phe Phe
 50 55 60
 Ser Phe Thr Ala Gly Lys Ala Pro His Lys Ser Leu Ser Val Met Leu
 65 70 75 80
 Val Arg Asn Arg Asp Glu Val Gln Ala Leu Ala Phe Asp Glu Gln Arg
 85 90 95
 Arg Pro Gly Ala Arg Arg Ala Ala Ser Gln Ser Ala Met Leu Gln Leu
 100 105 110
 Asp Tyr Gly Asp Thr Val Trp Leu Arg Leu His Gly Ala Pro Gln Tyr
 115 120 125
 Ala Leu Gly Ala Pro Gly Ala Thr Phe Ser Gly Tyr Leu Val Tyr Ala
 130 135 140
 Asp Ala Asp Ala Asp Ala Pro Ala Arg Gly Pro Pro Ala Pro Pro Glu
 145 150 155 160
 Pro Arg Ser Ala Phe Ser Ala Ala Arg Thr Arg Ser Leu Val Gly Ser
 165 170 175
 Asp Ala Gly Pro Gly Pro Arg His Gln Pro Leu Ala Phe Asp Thr Glu
 180 185 190
 Phe Val Asn Ile Gly Gly Asp Phe Asp Ala Ala Ala Gly Val Phe Arg
 195 200 205
 Cys Arg Leu Pro Gly Ala Tyr Phe Phe Ser Phe Thr Leu Gly Lys Leu
 210 215 220
 Pro Arg Lys Thr Leu Ser Val Lys Leu Met Lys Asn Arg Asp Glu Val
 225 230 235 240
 Gln Ala Met Ile Tyr Asp Asp Gly Ala Ser Arg Arg Arg Glu Met Gln
 245 250 255
 Ser Gln Ser Val Met Leu Ala Leu Arg Arg Gly Asp Ala Val Trp Leu
 260 265 270
 Leu Ser His Asp His Asp Gly Tyr Gly Ala Tyr Ser Asn His Asp Leu
 275 280 285
 Pro Thr Asp Leu Lys Thr Val Leu Pro Ser Trp Asp Val His Cys Cys
 290 295 300
 Gln Val Asn Gln Arg Phe Glu Leu Cys Ile Gly Val Ile Pro Glu Glu
 305 310 315 320
 Ser Gln His Trp Asp Asp Ala Ile Arg Met Asp Thr Asp Leu
 325 330

<210> 68
 <211> 17
 <212> PRT
 <213> Catarrhini

<400> 68

Met Leu Pro Leu Leu Leu Gly Leu Leu Gly Pro Ala Ala Cys Trp Ala
 1 5 10 15
 Leu

<210> 69
 <211> 130
 <212> PRT
 <213> Catarrhini

<400> 69
 Gly Pro Ala Pro Gly Pro Gly Ser Ser Glu Leu Arg Ser Ala Phe Ser
 1 5 10 15
 Ala Ala Arg Thr Thr Pro Leu Glu Gly Ala Ser Glu Met Ala Val Thr
 20 25 30
 Phe Asp Lys Val Tyr Val Asn Ile Gly Gly Asp Phe Asp Ala Ala Thr
 35 40 45
 Gly Gln Phe Arg Cys Arg Val Pro Gly Ala Tyr Phe Phe Ser Phe Thr
 50 55 60
 Val Gly Lys Ala Pro His Lys Ser Leu Ser Val Met Leu Ala Leu Arg
 65 70 75 80
 Arg Gly Asp Ala Val Trp Leu Leu Ser His Asp His Asp Gly Tyr Gly
 85 90 95
 Ala Tyr Ser Asn His Gly Lys Tyr Ile Thr Phe Ser Gly Phe Leu Val
 100 105 110
 Tyr Pro Asp Leu Ala Gly Gly Ala Pro Pro Gly Leu Gly Ala Pro Glu
 115 120 125
 Leu Leu
 130

<210> 70
 <211> 126
 <212> PRT
 <213> Catarrhini

<400> 70
 Ala Phe Ser Ala Ala Arg Thr Thr Pro Leu Glu Gly Thr Ser Glu Met
 1 5 10 15
 Ala Val Thr Phe Asp Lys Val Tyr Val Asn Ile Gly Gly Asp Phe Asp
 20 25 30
 Val Ala Thr Gly Gln Phe Arg Cys Arg Val Pro Gly Ala Tyr Phe Phe
 35 40 45
 Ser Phe Thr Ala Gly Lys Ala Pro His Lys Ser Leu Ser Val Met Leu
 50 55 60
 Val Arg Asn Arg Asp Glu Val Gln Ala Leu Ala Phe Asp Glu Gln Arg
 65 70 75 80
 Arg Pro Gly Ala Arg Arg Ala Ala Ser Gln Ser Ala Met Leu Gln Leu
 85 90 95
 Asp Tyr Gly Asp Thr Val Trp Leu Arg Leu His Gly Ala Pro Gln Tyr
 100 105 110
 Ala Leu Gly Ala Pro Gly Ala Thr Phe Ser Gly Tyr Leu Val
 115 120 125

<210> 71
 <211> 117
 <212> PRT
 <213> Catarrhini

<400> 71

```

Ala Phe Ser Ala Ala Arg Thr Arg Ser Leu Val Gly Ser Asp Ala Gly
 1             5             10             15
Pro Gly Pro Arg His Gln Pro Leu Ala Phe Asp Thr Glu Phe Val Asn
             20             25             30
Ile Gly Gly Asp Phe Asp Ala Ala Ala Gly Val Phe Arg Cys Arg Leu
             35             40             45
Pro Gly Ala Tyr Phe Phe Ser Phe Thr Leu Gly Lys Leu Pro Arg Lys
             50             55             60
Thr Leu Ser Val Lys Leu Met Lys Asn Arg Asp Glu Val Gln Ala Met
             65             70             75             80
Ile Tyr Asp Asp Gly Ala Ser Arg Arg Arg Glu Met Gln Ser Gln Ser
             85             90             95
Val Met Leu Ala Leu Arg Arg Gly Asp Ala Val Trp Leu Leu Ser His
             100             105             110
Asp His Asp Gly Tyr
             115

```

<210> 72

<211> 137

<212> PRT

<213> Artificial sequence

<220>

<223> Synthetically generated peptide

<400> 72

```

Ala Phe Thr Val Leu Arg Ser Thr Asn Arg Pro Pro Ala Glu Met Ser
 1             5             10             15
Asn Pro Gly Gln Pro Val Ile Phe Asp Glu Val Leu Tyr Asn Gln Gln
             20             25             30
Gly His Tyr Asp Pro Ala Thr Gly Lys Phe Thr Cys Lys Val Pro Gly
             35             40             45
Leu Tyr Tyr Phe Ser Phe His Val Ser Ser Lys Gly Thr Arg Gln Asn
             50             55             60
Val Cys Val Ser Leu Met Arg Ser Ser Arg Asn Gly Val Arg Gln Lys
             65             70             75             80
Val Met Glu Phe Cys Asp Glu Tyr Ala Lys Gly Thr Tyr Gln Val Ala
             85             90             95
Ser Gly Gly Ala Val Leu Gln Leu Arg Gln Gly Asp Arg Val Trp Leu
             100             105             110
Glu Leu Asp Asp Lys Gln Thr Asn Gly Leu Leu Gly Gly Glu Gly Val
             115             120             125
His Ser Val Phe Ser Gly Phe Leu Leu
             130             135

```

<210> 73

<211> 126

<212> PRT

<213> Catarrhini

<400> 73

```

Ala Phe Ser Ala Ala Arg Thr Thr Pro Leu Glu Gly Ala Ser Glu Met
 1             5             10             15
Ala Val Thr Phe Asp Lys Val Tyr Val Asn Ile Gly Gly Asp Phe Asp
             20             25             30
Ala Ala Thr Gly Gln Phe Arg Cys Arg Val Pro Gly Ala Tyr Phe Phe
             35             40             45

```

Ser Phe Thr Val Gly Lys Ala Pro His Lys Ser Leu Ser Val Met Leu
 50 55 60
 Val Arg Asn His Asp Glu Val Gln Ala Leu Ala Phe Asp Glu Gln Arg
 65 70 75 80
 Arg Pro Ser Ala Arg Arg Ala Ala Ser Gln Ser Ala Met Leu Gln Leu
 85 90 95
 Asp Tyr Gly Asp Thr Val Trp Leu Arg Leu His Gly Ala Pro Gln Tyr
 100 105 110
 Ala Leu Gly Ala Pro Gly Ala Thr Phe Ser Gly Tyr Leu Val
 115 120 125

<210> 74
 <211> 134
 <212> PRT
 <213> Catarrhini

<400> 74
 Ala Phe Ser Ala Ala Arg Thr Arg Ser Leu Val Gly Ser Asp Ala Gly
 1 5 10 15
 Ser Gly Pro Arg His Arg Pro Leu Ala Phe Asp Thr Glu Leu Val Asn
 20 25 30
 Ile Gly Gly Asp Phe Asp Ala Ala Ala Gly Val Phe Arg Cys Arg Leu
 35 40 45
 Pro Gly Ala Tyr Phe Phe Ser Phe Thr Leu Gly Lys Leu Pro Arg Lys
 50 55 60
 Thr Leu Ser Val Lys Leu Met Lys Asn Arg Asp Glu Val Gln Ala Met
 65 70 75 80
 Ile Tyr Asp Asp Gly Ala Ser Arg Arg Arg Glu Met Gln Ser Gln Ser
 85 90 95
 Val Met Leu Ala Leu Arg Arg Gly Asp Ala Val Trp Leu Leu Ser His
 100 105 110
 Asp His Asp Gly Tyr Gly Ala Tyr Ser Asn His Gly Lys Tyr Ile Thr
 115 120 125
 Phe Ser Gly Phe Leu Val
 130

<210> 75
 <211> 93
 <212> PRT
 <213> Homo sapiens

<400> 75
 Leu Cys Ser Gln Ser Gly Gln Thr Ser Val Gly Gly Ser Thr Ala Leu
 1 5 10 15
 Arg Cys Ser Ser Ser Glu Gly Ala Pro Lys Pro Val Tyr Asn Trp Val
 20 25 30
 Arg Leu Gly Thr Phe Pro Thr Pro Ser Pro Gly Ser Met Val Gln Asp
 35 40 45
 Glu Val Ser Gly Gln Leu Ile Leu Thr Asn Leu Ser Leu Thr Ser Ser
 50 55 60
 Gly Thr Tyr Arg Cys Val Ala Thr Asn Gln Met Gly Ser Ala Ser Cys
 65 70 75 80
 Glu Leu Thr Leu Ser Val Thr Glu Pro Ser Gln Gly Arg
 85 90

<210> 76
 <211> 23
 <212> PRT

<213> Homo sapiens

<400> 76

Val Ala Gly Ala Leu Ile Gly Val Leu Leu Gly Val Leu Leu Leu Ser
 1 5 10 15
 Val Ala Ala Phe Cys Leu Val
 20

<210> 77

<211> 20

<212> PRT

<213> Homo sapiens

<400> 77

Arg Phe Gln Lys Glu Arg Gly Lys Lys Pro Lys Glu Thr Tyr Gly Gly
 1 5 10 15
 Ser Asp Leu Arg
 20

<210> 78

<211> 125

<212> PRT

<213> Homo sapiens

<400> 78

Trp Ala Leu Gly Pro Thr Pro Gly Pro Gly Ser Ser Glu Leu Arg Ser
 1 5 10 15
 Ala Phe Ser Ala Ala Arg Thr Thr Pro Leu Glu Gly Thr Ser Glu Met
 20 25 30
 Ala Val Thr Phe Asp Lys Val Tyr Val Asn Ile Gly Gly Asp Phe Asp
 35 40 45
 Val Ala Thr Gly Gln Phe Arg Cys Arg Val Pro Gly Ala Tyr Phe Phe
 50 55 60
 Ser Phe Thr Ala Gly Lys Ala Pro His Lys Ser Leu Ser Val Met Leu
 65 70 75 80
 Val Arg Asn Arg Asp Glu Val Gln Ala Leu Ala Phe Asp Glu Gln Arg
 85 90 95
 Arg Pro Gly Ala Arg Arg Ala Ala Ser Gln Ser Ala Met Leu Gln Leu
 100 105 110
 Asp Tyr Gly Asp Thr Val Trp Leu Arg Leu His Gly Ala
 115 120 125

<210> 79

<211> 17

<212> PRT

<213> Homo sapiens

<400> 79

Pro Gln Tyr Ala Leu Gly Ala Pro Gly Ala Thr Phe Ser Gly Tyr Leu
 1 5 10 15
 Val

<210> 80

<211> 192

<212> PRT

<213> Homo sapiens

<400> 80

```

Tyr Ala Asp Ala Asp Ala Asp Ala Pro Ala Arg Gly Pro Pro Ala Pro
 1           5           10           15
Pro Glu Pro Arg Ser Ala Phe Ser Ala Ala Arg Thr Arg Ser Leu Val
          20           25           30
Gly Ser Asp Ala Gly Pro Gly Pro Arg His Gln Pro Leu Ala Phe Asp
          35           40           45
Thr Glu Phe Val Asn Ile Gly Gly Asp Phe Asp Ala Ala Ala Gly Val
          50           55           60
Phe Arg Cys Arg Leu Pro Gly Ala Tyr Phe Phe Ser Phe Thr Leu Gly
 65           70           75           80
Lys Leu Pro Arg Lys Thr Leu Ser Val Lys Leu Met Lys Asn Arg Asp
          85           90           95
Glu Val Gln Ala Met Ile Tyr Asp Asp Gly Ala Ser Arg Arg Arg Glu
          100          105          110
Met Gln Ser Gln Ser Val Met Leu Ala Leu Arg Arg Gly Asp Ala Val
          115          120          125
Trp Leu Leu Ser His Asp His Asp Gly Tyr Gly Ala Tyr Ser Asn His
          130          135          140
Asp Leu Pro Thr Asp Leu Lys Thr Val Leu Pro Ser Trp Asp Val His
          145          150          155          160
Cys Cys Gln Val Asn Gln Arg Phe Glu Leu Cys Ile Gly Val Ile Pro
          165          170          175
Glu Glu Ser Gln His Trp Asp Asp Ala Ile Arg Met Asp Thr Asp Leu
          180          185          190

```

<210> 81

<211> 325

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> 20

<223> Xaa = Unknown amino acid

<400> 81

```

Met Ala Asp Leu Pro Gly Pro Phe Leu Cys Gly Ala Leu Leu Gly Phe
 1           5           10           15
Leu Cys Leu Xaa Leu Ala Val Glu Val Lys Val Pro Thr Glu Pro Leu
          20           25           30
Ser Thr Pro Leu Gly Lys Thr Ala Glu Leu Thr Cys Thr Tyr Ser Thr
          35           40           45
Ser Val Gly Asp Thr Phe Ala Leu Glu Trp Ser Phe Val Gln Pro Gly
          50           55           60
Lys Pro Ile Ser Glu Ser His Pro Ile Leu Tyr Phe Thr Asn Gly His
 65           70           75           80
Leu Tyr Pro Thr Gly Ser Lys Ser Lys Arg Val Ser Leu Leu Gln Asn
          85           90           95
Pro Pro Thr Val Gly Val Ala Thr Leu Lys Leu Thr Asp Val His Pro
          100          105          110
Ser Asp Thr Gly Thr Tyr Leu Cys Gln Val Asn Asn Pro Pro Asp Phe
          115          120          125
Tyr Thr Asn Gly Leu Gly Leu Ile Asn Leu Thr Val Leu Val Pro Pro
          130          135          140
Ser Asn Pro Leu Cys Ser Gln Ser Gly Gln Thr Ser Val Gly Gly Ser
          145          150          155          160

```

Thr Ala Leu Arg Cys Ser Ser Ser Glu Gly Ala Pro Lys Pro Val Tyr
 165 170 175
 Asn Trp Val Arg Leu Gly Thr Phe Pro Thr Pro Ser Pro Gly Ser Met
 180 185 190
 Val Gln Asp Glu Val Ser Gly Gln Leu Ile Leu Thr Asn Leu Ser Leu
 195 200 205
 Thr Ser Ser Gly Thr Tyr Arg Cys Val Ala Thr Asn Gln Leu Gly Ser
 210 215 220
 Ala Ser Cys Glu Leu Thr Leu Ser Val Thr Glu Pro Ser Gln Gly Arg
 225 230 235 240
 Val Thr Gly Ala Leu Ile Gly Val Leu Leu Gly Val Leu Leu Leu Ser
 245 250 255
 Val Ala Ala Phe Cys Leu Val Arg Phe Gln Lys Glu Arg Gly Lys Lys
 260 265 270
 Pro Lys Glu Thr Tyr Gly Gly Ser Asp Leu Arg Glu Asp Ala Ile Ala
 275 280 285
 Pro Gly Ile Ser Glu His Thr Cys Met Arg Ala Asp Ser Ser Lys Gly
 290 295 300
 Phe Leu Glu Arg Pro Ser Ala Ser Thr Val Thr Thr Thr Lys Ser Lys
 305 310 315 320
 Leu Pro Met Val Val
 325

<210> 82

<211> 4317

<212> DNA

<213> Mus musculus

<400> 82

cctcctgctg	tcccctccct	acccttggt	tctcgccccg	ctctgcectc	tgetaccaac	60
actcgatccc	ctgctcgggc	tcgacctcca	atctccgagg	gtcgtgcggc	cccggatgcc	120
cgggccccga	gcggtgccca	cggcctggcc	cctgcgatgc	gccaggccgg	cggattgggg	180
ctgctggcac	tactcctgct	ggcgctgctg	ggccccggcg	gccgaggggt	gggcccggcg	240
ggcagcgggg	cacaggcggg	ggcggggcgc	tgggcccac	gcttcaaggt	ggtctttgcg	300
cctgtgatct	gcaagcggac	ctgtctgaag	ggccagtgtc	gggacagctg	tcagcagggc	360
tccaacatga	cgctcatcgg	agagaacggc	cacagcaccg	acacgctcac	cggttctgcc	420
ttccgcgtgg	tgggtgtgcc	tctaccctgc	atgaacggtg	gccagtgtct	ttcccgaaac	480
cagtgcctgt	gtcccccgga	tttcacgggg	cgcttctgcc	aggtgcctgc	tgcaggaacc	540
ggagctggca	ccgggagttc	aggccccggc	tggcccgacc	gggccatgtc	cacaggcccg	600
ctgccgcccc	ttgccccaga	aggagagtct	gtggctagca	aacacgccat	ttacgcggtg	660
caggtgatcg	cagatcctcc	cgggccgggg	gagggctctc	ctgcacaaca	tgcagccttc	720
ttggtgcccc	tggggccagg	acaaatctcg	gcagaagtgc	aggctccgcc	ccccgtggtg	780
aacgtgctg	tccatcacc	tctgaagct	tccgttcagg	tgcaccgcat	cgaggggccc	840
aacgtgtaag	gccagcctc	ttcccagcac	ttgctgccgc	atcccaagcc	cccgcccccg	900
aggccaccca	ctcaaaagcc	actgggccgc	tgcttccagg	acacattgcc	caagcagcct	960
tgtggcagca	accctttgcc	tggccttacc	aagcaggaag	attgctgcgg	tagcatcggt	1020
actgcctggg	gacaaagcaa	gtgtcacaa	tgcccacagc	ttcagtatac	aggggtgcag	1080
aagcctgtac	ctgtacgtgg	ggaggtgggt	gctgactgcc	cccagggtta	caagaggtct	1140
aacagcacc	actgccagga	tatcaacgaa	tgtgcgatgc	ccgggaatgt	gtgccatggt	1200
gactgcctca	acaacctg	ctcttatcgc	tgtgtctgcc	cgcccgggtc	tagcttgggt	1260
ccccctcgag	cacagtgc	tgcgacaaa	ccagaggaga	agagcctgtg	ttccgcctt	1320
gtgagcaccg	aacaccagt	ccagcaccct	ctgaccacac	gcctaaccgg	ccagctctgc	1380
tgctgtagtg	tgggtaaagc	ctgggtgccc	cggtgccagc	gctgcccggc	agatgtgtaca	1440
gcagccttca	aggagatctg	ccccggctgg	gaaaggggtac	catatcctca	cctcccacca	1500
gacgtcacc	atccagggg	aaagcgactt	ctccctcttc	ctgcaccgca	cggggccacc	1560
aaaccccagc	agcttctctga	aagccccagc	cgagcaccac	ccctcgagga	cacagaggaa	1620
gagagaggag	tgaccatgga	tccaccagt	agtgaggagc	gatcggtgca	gcagagccac	1680
cccactacca	ccacctcacc	ccccggcct	taccagagc	tcattctctg	cccccccca	1740

```

cctaccttcc accggttcct gccagacttg ccccatcccc gaagtgcagt ggagatcgcc 1800
cccatcagg tcacagagac cgatgagtgc cgattgaacc agaatatctg tggccatgga 1860
cagtgtgtgc ctggcccctc ggattactcc tgccactgca acgctggcta ccggtcacac 1920
ccgcagcacc gctactgtgt tgatgtgaac gagtgcgagg cagagccctg cggccccggg 1980
aaaggcatct gtatgaacac tgggtggctcc tacaattgtc actgcaaccg aggctaccgc 2040
ctccacgtgg gtgcaggggg ccgctcgtgc gtggacctga acgagtgcgc caagcctcac 2100
ctgtgtgggg acggtggctt ctgcatcaac ttccctggtc actacaaatg caactgctat 2160
cctggctacc ggctcaaggc ctcccagacc cccatttgcg aagacatcga cgcgtgtcgc 2220
gaccctagca cctgccctga tggcaaatgt gaaaacaaac ctggcagctt caagtgcac 2280
gcttgccagc ctggctaccg tagccagggg ggcggggcct gtcgtgatgt caacgaatgc 2340
tccgagggta cccctgtctc tcctggatgg tgtgagaacc ttccgggttc ttaccgttgc 2400
acgtgtgccc aggggatacg aaccgcaca ggacgcctca gttgcataga cgtggatgag 2460
tgtgaggctg ggaaagtgtg ccaagatggc atctgcacga acacaccagg ctctttccag 2520
tgtcagtgcc tctccggcta tcatctgtca agggatcgga gccgctgtga ggacattgat 2580
gaatgtgact tccctgcggc ctgcatcggg ggtgactgca tcaataccaa tggttcctac 2640
agatgtctct gtccctggg tcatcggttg gtggcgggca ggaagtgcaa gaaagatata 2700
gatgagtgca gccaggaccc aggctgtgct gtgccccatg cctgcgagaa cctccagggc 2760
tcctatgtct gtgtctgtga tgagggtttc acactcacc aggaccagca tgggtgtgag 2820
gaggtggagc agccccacca caagaaggag tgctacctta acttcgatga cacagtgttc 2880
tgtgacagcg tattggctac caatgtcact cagcaggaat gctgttgctc tctgggagct 2940
ggctggggag accactgcga aatctatccc tgtccagtct acagctcagc cgaatttcac 3000
agcctggtgc ctgatgggaa aaggctacac tcaggacaac aacattgtga actatgcatt 3060
cctgccacc gtgacatcga cgaatgcata ttgtttgggg cagagatctg caaggagggc 3120
aagtgtgtga acacgcagcc cggctacgag tgctactgca agcagggctt ctactacgat 3180
ggcaacctgc tggagtgcgt ggacgtggat gagtgccttg atgagtctaa ctgcaggaa 3240
ggagtgtgtg agaacacacg tggcggttac cgctgtgcct gcactccgcc ggcagagtac 3300
agtcccgcac aggccagtg tctgatcccg gagagatgga gcacgcccc gagagacgtg 3360
aagtgtgctg gggccagcga ggagaggacg gcatgtgtat ggggcccctg ggcgggacct 3420
gccctcactt ttgatgactg ctgctgccgc cagccgcggc tgggtaccca gtgcagaccg 3480
tgcccgcac gtggcaccgg gtcccagtg cgaattcac agagtgaag caattctttc 3540
tgggacacaa gccccctgct actggggaag tctccgcgag acgaagacag ctgagaggag 3600
gattcagatg agtgccgttg tgtgagcgga cgctgtgtgc cagggccagg cggggcggtg 3660
tgcgagtgtc ctggaggctt tcagctggac gcctccgtg cccgctgcgt ggacattgat 3720
gagtgcgag aactgaacca gcgggagctg ctgtgtaaga gcgagcgggtg cgtgaacacc 3780
agtggatcct tccgctgtgt ctgcaaagct ggccttcacgc gcagccgcc tcacgggcct 3840
gcgtgcctca gcgcgcgc tgatgatgca gccatagccc acacctcagt gatcgatcat 3900
cgagggtatt ttactgaaa gtggagacag acaagtacat cctttgctcc tgaccaaaccg 3960
agagcatgga ccaaggatc cttcagggcc cacaatctc cttccacac cccaaaccca 4020
agggtgctct gtctgcagag tgctgtctgc ttttcccaa gggtgattcc tagaaacttc 4080
gacatcagat ctgccccttt aatttactct tggctttcaa ggcaaattga tattcacatc 4140
caaagcgggc agcatcaact gcttggcggg ttggactgag ctgggaccca ggatgtgaaa 4200
tagaatttat tgtggctctg attatgtaca ctgatgtgc ctgacctgct gaccaggctc 4260
acatggtttg tacaataaat acatccgccg ggaaaaaaaa aaaaaaaaaa aaaaaaa 4317

```

<210> 83

<211> 1251

<212> PRT

<213> Mus musculus

<400> 83

```

Met Arg Gln Ala Ala Leu Gly Leu Leu Ala Leu Leu Leu Ala Leu
 1               5               10              15
Leu Gly Pro Gly Gly Arg Gly Val Gly Arg Pro Gly Ser Gly Ala Gln
      20              25              30
Ala Gly Ala Gly Arg Trp Ala Gln Arg Phe Lys Val Val Phe Ala Pro
      35              40              45
Val Ile Cys Lys Arg Thr Cys Leu Lys Gly Gln Cys Arg Asp Ser Cys
      50              55              60

```

Gln	Gln	Gly	Ser	Asn	Met	Thr	Leu	Ile	Gly	Glu	Asn	Gly	His	Ser	Thr	
65					70					75					80	
Asp	Thr	Leu	Thr	Gly	Ser	Ala	Phe	Arg	Val	Val	Val	Cys	Pro	Leu	Pro	
				85					90					95		
Cys	Met	Asn	Gly	Gly	Gln	Cys	Ser	Ser	Arg	Asn	Gln	Cys	Leu	Cys	Pro	
			100					105					110			
Pro	Asp	Phe	Thr	Gly	Arg	Phe	Cys	Gln	Val	Pro	Ala	Ala	Gly	Thr	Gly	
		115					120					125				
Ala	Gly	Thr	Gly	Ser	Ser	Gly	Pro	Gly	Trp	Pro	Asp	Arg	Ala	Met	Ser	
	130					135					140					
Thr	Gly	Pro	Leu	Pro	Pro	Leu	Ala	Pro	Glu	Gly	Glu	Ser	Val	Ala	Ser	
145					150					155					160	
Lys	His	Ala	Ile	Tyr	Ala	Val	Gln	Val	Ile	Ala	Asp	Pro	Pro	Gly	Pro	
			165					170						175		
Gly	Glu	Gly	Pro	Pro	Ala	Gln	His	Ala	Ala	Phe	Leu	Val	Pro	Leu	Gly	
		180						185					190			
Pro	Gly	Gln	Ile	Ser	Ala	Glu	Val	Gln	Ala	Pro	Pro	Pro	Val	Val	Asn	
		195					200					205				
Val	Arg	Val	His	His	Pro	Pro	Glu	Ala	Ser	Val	Gln	Val	His	Arg	Ile	
	210					215					220					
Glu	Gly	Pro	Asn	Ala	Glu	Gly	Pro	Ala	Ser	Ser	Gln	His	Leu	Leu	Pro	
225				230						235					240	
His	Pro	Lys	Pro	Gln	His	Pro	Arg	Pro	Pro	Thr	Gln	Lys	Pro	Leu	Glu	
				245				250						255		
Arg	Cys	Phe	Gln	Asp	Thr	Leu	Pro	Lys	Gln	Pro	Cys	Gly	Ser	Asn	Pro	
		260						265					270			
Leu	Pro	Gly	Leu	Thr	Lys	Gln	Glu	Asp	Cys	Cys	Gly	Ser	Ile	Gly	Thr	
		275					280					285				
Ala	Trp	Gly	Gln	Ser	Lys	Cys	His	Lys	Cys	Pro	Gln	Leu	Gln	Tyr	Thr	
	290					295					300					
Gly	Val	Gln	Lys	Pro	Val	Pro	Val	Arg	Gly	Glu	Val	Gly	Ala	Asp	Cys	
305				310						315					320	
Pro	Gln	Gly	Tyr	Lys	Arg	Leu	Asn	Ser	Thr	His	Cys	Gln	Asp	Ile	Asn	
			325					330						335		
Glu	Cys	Ala	Met	Pro	Gly	Asn	Val	Cys	His	Gly	Asp	Cys	Leu	Asn	Asn	
		340						345					350			
Pro	Gly	Ser	Tyr	Arg	Cys	Val	Cys	Pro	Pro	Gly	His	Ser	Leu	Gly	Pro	
		355					360					365				
Leu	Ala	Ala	Gln	Cys	Ile	Ala	Asp	Lys	Pro	Glu	Glu	Lys	Ser	Leu	Cys	
	370					375					380					
Phe	Arg	Leu	Val	Ser	Thr	Glu	His	Gln	Cys	Gln	His	Pro	Leu	Thr	Thr	
385					390					395					400	
Arg	Leu	Thr	Arg	Gln	Leu	Cys	Cys	Cys	Ser	Val	Gly	Lys	Ala	Trp	Gly	
				405				410						415		
Ala	Arg	Cys	Gln	Arg	Cys	Pro	Ala	Asp	Gly	Thr	Ala	Ala	Phe	Lys	Glu	
			420					425					430			
Ile	Cys	Pro	Gly	Trp	Glu	Arg	Val	Pro	Tyr	Pro	His	Leu	Pro	Pro	Asp	
	435						440					445				
Ala	His	His	Pro	Gly	Gly	Lys	Arg	Leu	Leu	Pro	Leu	Pro	Ala	Pro	Asp	
	450					455					460					
Gly	Pro	Pro	Lys	Pro	Gln	Gln	Leu	Pro	Glu	Ser	Pro	Ser	Arg	Ala	Pro	
465					470					475					480	
Pro	Leu	Glu	Asp	Thr	Glu	Glu	Glu	Arg	Gly	Val	Thr	Met	Asp	Pro	Pro	
			485					490						495		
Val	Ser	Glu	Glu	Arg	Ser	Val	Gln	Gln	Ser	His	Pro	Thr	Thr	Thr	Thr	
			500					505					510			
Ser	Pro	Pro	Arg	Pro	Tyr	Pro	Glu	Leu	Ile	Ser	Arg	Pro	Ser	Pro	Pro	
		515					520					525				

Thr	Phe	His	Arg	Phe	Leu	Pro	Asp	Leu	Pro	Pro	Ser	Arg	Ser	Ala	Val
530						535					540				
Glu	Ile	Ala	Pro	Thr	Gln	Val	Thr	Glu	Thr	Asp	Glu	Cys	Arg	Leu	Asn
545					550					555					560
Gln	Asn	Ile	Cys	Gly	His	Gly	Gln	Cys	Val	Pro	Gly	Pro	Ser	Asp	Tyr
				565					570					575	
Ser	Cys	His	Cys	Asn	Ala	Gly	Tyr	Arg	Ser	His	Pro	Gln	His	Arg	Tyr
				580				585					590		
Cys	Val	Asp	Val	Asn	Glu	Cys	Glu	Ala	Glu	Pro	Cys	Gly	Pro	Gly	Lys
		595					600					605			
Gly	Ile	Cys	Met	Asn	Thr	Gly	Gly	Ser	Tyr	Asn	Cys	His	Cys	Asn	Arg
610						615				620					
Gly	Tyr	Arg	Leu	His	Val	Gly	Ala	Gly	Gly	Arg	Ser	Cys	Val	Asp	Leu
625					630					635					640
Asn	Glu	Cys	Ala	Lys	Pro	His	Leu	Cys	Gly	Asp	Gly	Gly	Phe	Cys	Ile
				645					650					655	
Asn	Phe	Pro	Gly	His	Tyr	Lys	Cys	Asn	Cys	Tyr	Pro	Gly	Tyr	Arg	Leu
			660					665					670		
Lys	Ala	Ser	Arg	Pro	Pro	Ile	Cys	Glu	Asp	Ile	Asp	Glu	Cys	Arg	Asp
			675				680					685			
Pro	Ser	Thr	Cys	Pro	Asp	Gly	Lys	Cys	Glu	Asn	Lys	Pro	Gly	Ser	Phe
		690				695					700				
Lys	Cys	Ile	Ala	Cys	Gln	Pro	Gly	Tyr	Arg	Ser	Gln	Gly	Gly	Gly	Ala
705					710					715					720
Cys	Arg	Asp	Val	Asn	Glu	Cys	Ser	Glu	Gly	Thr	Pro	Cys	Ser	Pro	Gly
				725					730					735	
Trp	Cys	Glu	Lys	Leu	Pro	Gly	Ser	Tyr	Arg	Cys	Thr	Cys	Ala	Gln	Gly
			740					745					750		
Ile	Arg	Thr	Arg	Thr	Gly	Arg	Leu	Ser	Cys	Ile	Asp	Val	Asp	Asp	Cys
		755					760					765			
Glu	Ala	Gly	Lys	Val	Cys	Gln	Asp	Gly	Ile	Cys	Thr	Asn	Thr	Pro	Gly
		770				775					780				
Ser	Phe	Gln	Cys	Gln	Cys	Leu	Ser	Gly	Tyr	His	Leu	Ser	Arg	Asp	Arg
785					790					795					800
Ser	Arg	Cys	Glu	Asp	Ile	Asp	Glu	Cys	Asp	Phe	Pro	Ala	Ala	Cys	Ile
				805					810					815	
Gly	Gly	Asp	Cys	Ile	Asn	Thr	Asn	Gly	Ser	Tyr	Arg	Cys	Leu	Cys	Pro
			820					825					830		
Leu	Gly	His	Arg	Leu	Val	Gly	Gly	Arg	Lys	Cys	Lys	Lys	Asp	Ile	Asp
		835					840					845			
Glu	Cys	Ser	Gln	Asp	Pro	Gly	Leu	Cys	Leu	Pro	His	Ala	Cys	Glu	Asn
		850				855				860					
Leu	Gln	Gly	Ser	Tyr	Val	Cys	Val	Cys	Asp	Glu	Gly	Phe	Thr	Leu	Thr
865					870					875					880
Gln	Asp	Gln	His	Gly	Cys	Glu	Glu	Val	Glu	Gln	Pro	His	His	Lys	Lys
				885					890					895	
Glu	Cys	Tyr	Leu	Asn	Phe	Asp	Asp	Thr	Val	Phe	Cys	Asp	Ser	Val	Leu
			900					905					910		
Ala	Thr	Asn	Val	Thr	Gln	Gln	Glu	Cys	Cys	Cys	Ser	Leu	Gly	Ala	Gly
		915					920					925			
Trp	Gly	Asp	His	Cys	Glu	Ile	Tyr	Pro	Cys	Pro	Val	Tyr	Ser	Ser	Ala
		930				935					940				
Glu	Phe	His	Ser	Leu	Val	Pro	Asp	Gly	Lys	Arg	Leu	His	Ser	Gly	Gln
945					950					955					960
Gln	His	Cys	Glu	Leu	Cys	Ile	Pro	Ala	His	Arg	Asp	Ile	Asp	Glu	Cys
				965					970					975	
Ile	Leu	Phe	Gly	Ala	Glu	Ile	Cys	Lys	Glu	Gly	Lys	Cys	Val	Asn	Ser
			980					985					990		

Gln Pro Gly Tyr Glu Cys Tyr Cys Lys Gln Gly Phe Tyr Tyr Asp Gly
 995 1000 1005
 Asn Leu Leu Glu Cys Val Asp Val Asp Glu Cys Leu Asp Glu Ser Asn
 1010 1015 1020
 Cys Arg Asn Gly Val Cys Glu Asn Thr Trp Arg Leu Pro Cys Ala Cys
 1025 1030 1035 1040
 Thr Pro Pro Ala Glu Tyr Ser Pro Ala Gln Ala Gln Cys Leu Ser Pro
 1045 1050 1055
 Glu Glu Met Glu His Ala Pro Glu Arg Arg Glu Val Cys Trp Gly Gln
 1060 1065 1070
 Arg Gly Glu Asp Gly Met Cys Met Gly Pro Leu Ala Gly Pro Ala Leu
 1075 1080 1085
 Thr Phe Asp Asp Cys Cys Cys Arg Gln Pro Arg Leu Gly Tyr Gln Cys
 1090 1095 1100
 Arg Pro Cys Pro Pro Arg Gly Thr Gly Ser Gln Cys Pro Thr Ser Gln
 1110 1115 1120
 Ser Glu Ser Asn Ser Phe Trp Asp Thr Ser Pro Leu Leu Leu Gly Lys
 1125 1130 1135
 Ser Pro Arg Asp Glu Asp Ser Ser Glu Glu Asp Ser Asp Glu Cys Arg
 1140 1145 1150
 Cys Val Ser Gly Pro Cys Val Pro Arg Pro Gly Gly Ala Val Cys Glu
 1155 1160 1165
 Cys Pro Gly Gly Phe Gln Leu Asp Ala Ser Arg Ala Arg Cys Val Asp
 1170 1175 1180
 Ile Asp Glu Cys Arg Glu Leu Asn Gln Arg Gly Leu Leu Cys Lys Ser
 1190 1195 1200
 Glu Arg Cys Val Asn Thr Ser Gly Ser Phe Arg Cys Val Cys Lys Ala
 1205 1210 1215
 Gly Phe Thr Arg Ser Arg Pro His Gly Pro Ala Cys Leu Ser Ala Ala
 1220 1225 1230
 Ala Asp Asp Ala Ala Ile Ala His Thr Ser Val Ile Asp His Arg Gly
 1235 1240 1245
 Tyr Phe His
 1250

<210> 84
 <211> 353
 <212> PRT
 <213> Arabidopsis thaliana

<400> 84
 Met Ala Asn Asp Gln Lys Asn Ser Glu Ser Phe Pro Ala Lys Glu Asp
 1 5 10 15
 His Lys Lys Asp Asp Ala Ala Ala Pro Ala Glu Val Asp His Lys Asp
 20 25 30
 Glu Phe Ser Ala Ser Gln Pro His Pro Val Glu Asn Ile Val Leu Val
 35 40 45
 Gly Arg Thr Gly Asn Gly Lys Ser Ala Thr Gly Asn Ser Ile Val Arg
 50 55 60
 Ser Lys Val Phe Lys Ser Lys Thr Lys Ser Ser Gly Val Thr Met Glu
 65 70 75 80
 Cys His Ala Val Lys Ala Val Thr Pro Glu Gly Pro Ile Leu Asn Val
 85 90 95
 Ile Asp Thr Pro Gly Leu Phe Asp Leu Ser Val Ser Ala Glu Phe Ile
 100 105 110
 Gly Lys Glu Ile Val Lys Cys Leu Thr Leu Ala Asp Gly Gly Leu His
 115 120 125

Ala Val Leu Leu Val Leu Ser Val Arg Thr Arg Ile Ser Gln Glu Glu
 130 135 140
 Glu Met Val Leu Ser Thr Leu Gln Val Leu Phe Gly Ser Lys Ile Val
 145 150 155 160
 Asp Tyr Leu Ile Val Val Phe Thr Gly Gly Asp Val Leu Glu Asp Asp
 165 170 175
 Gly Met Thr Leu Glu Asp Tyr Leu Gly Asp Asn Met Pro Asp Phe Leu
 180 185 190
 Lys Arg Val Leu Ile Leu Cys Gly Gln Arg Met Ile Leu Phe Asp Asn
 195 200 205
 Lys Thr Lys Asp Asp Glu Lys Lys Thr Lys Gln Val His Glu Leu Leu
 210 215 220
 Lys Leu Ile Asp Leu Val Arg Lys Gln Asn Asn Ile Pro Tyr Thr
 225 230 235 240
 Asp Glu Met Tyr His Met Ile Lys Glu Glu Asn Glu Arg His Lys Lys
 245 250 255
 Glu Gln Glu Glu Leu Glu Ser Lys Gly His Ser Glu Glu Gln Leu Ala
 260 265 270
 Ala Leu Met Lys Glu Leu Gln Ile Met Asn Glu Arg Asn Leu Lys Ala
 275 280 285
 Met Ala Glu Met Met Glu Lys Asn Met Lys Ile Ala Met Glu Ala Gln
 290 295 300
 Glu Lys Leu Phe Glu Gln Arg Glu Lys Ala Gln Glu Met Ser Tyr Gln
 305 310 315 320
 Gln Lys Met Glu Met Gln Glu Lys Leu Lys Gln Met Glu Gly Arg Met
 325 330 335
 Arg Ala Glu Met Glu Ala Gln Met Leu Ser Arg Gln Gln Cys Ser Ile
 340 345 350
 Leu

<210> 85

<211> 1251

<212> PRT

<213> Homo sapiens

<400> 85

Met Arg Gln Ala Ala Leu Gly Leu Leu Ala Leu Leu Leu Leu Ala Leu
 1 5 10 15
 Leu Gly Pro Gly Gly Arg Gly Val Gly Arg Pro Gly Ser Gly Ala Gln
 20 25 30
 Ala Gly Ala Gly Arg Trp Ala Gln Arg Phe Lys Val Val Phe Ala Pro
 35 40 45
 Val Ile Cys Lys Arg Thr Cys Leu Lys Gly Gln Cys Arg Asp Ser Cys
 50 55 60
 Gln Gln Gly Ser Asn Met Thr Leu Ile Gly Glu Asn Gly His Ser Thr
 65 70 75 80
 Asp Thr Leu Thr Gly Ser Ala Phe Arg Val Val Val Cys Pro Leu Pro
 85 90 95
 Cys Met Asn Gly Gly Gln Cys Ser Ser Arg Asn Gln Cys Leu Cys Pro
 100 105 110
 Pro Asp Phe Thr Gly Arg Phe Cys Gln Val Pro Ala Ala Gly Thr Gly
 115 120 125
 Ala Gly Thr Gly Ser Ser Gly Pro Gly Trp Pro Asp Arg Ala Met Ser
 130 135 140
 Thr Gly Pro Leu Pro Pro Leu Ala Pro Glu Gly Glu Ser Val Ala Ser
 145 150 155 160

Lys His Ala Ile Tyr Ala Val Gln Val Ile Ala Asp Pro Pro Gly Pro
 165 170 175
 Gly Glu Gly Pro Pro Ala Gln His Ala Ala Phe Leu Val Pro Leu Gly
 180 185 190
 Pro Gly Gln Ile Ser Ala Glu Val Gln Ala Pro Pro Pro Val Val Asn
 195 200 205
 Val Arg Val His His Pro Pro Glu Ala Ser Val Gln Val His Arg Ile
 210 215 220
 Glu Gly Pro Asn Ala Glu Gly Pro Ala Ser Ser Gln His Leu Leu Pro
 225 230 235 240
 His Pro Lys Pro Gln His Pro Arg Pro Pro Thr Gln Lys Pro Leu Gly
 245 250 255
 Arg Cys Phe Gln Asp Thr Leu Pro Lys Gln Pro Cys Gly Ser Asn Pro
 260 265 270
 Leu Pro Gly Leu Thr Lys Gln Glu Asp Cys Cys Gly Ser Ile Gly Thr
 275 280 285
 Ala Trp Gly Gln Ser Lys Cys His Lys Cys Pro Gln Leu Gln Tyr Thr
 290 295 300
 Gly Val Gln Lys Pro Val Pro Val Arg Gly Glu Val Gly Ala Asp Cys
 305 310 315 320
 Pro Gln Gly Tyr Lys Arg Leu Asn Ser Thr His Cys Gln Asp Ile Asn
 325 330 335
 Glu Cys Ala Met Pro Gly Asn Val Cys His Gly Asp Cys Leu Asn Asn
 340 345 350
 Pro Gly Ser Tyr Arg Cys Val Cys Pro Pro Gly His Ser Leu Gly Pro
 355 360 365
 Leu Ala Ala Gln Cys Ile Ala Asp Lys Pro Glu Glu Lys Ser Leu Cys
 370 375 380
 Phe Arg Leu Val Ser Thr Glu His Gln Cys Gln His Pro Leu Thr Thr
 385 390 395 400
 Arg Leu Thr Arg Gln Leu Cys Cys Cys Ser Val Gly Lys Ala Trp Gly
 405 410 415
 Ala Arg Cys Gln Arg Cys Pro Ala Asp Gly Thr Ala Ala Phe Lys Glu
 420 425 430
 Ile Cys Pro Gly Trp Glu Arg Val Pro Tyr Pro His Leu Pro Pro Asp
 435 440 445
 Ala His His Pro Gly Gly Lys Arg Leu Leu Pro Leu Pro Ala Pro Asp
 450 455 460
 Gly Pro Pro Lys Pro Gln Gln Leu Pro Glu Ser Pro Ser Arg Ala Pro
 465 470 475 480
 Pro Leu Glu Asp Thr Glu Glu Glu Arg Gly Val Thr Met Asp Pro Pro
 485 490 495
 Val Ser Glu Glu Arg Ser Val Gln Gln Ser His Pro Thr Thr Thr Thr
 500 505 510
 Ser Pro Pro Arg Pro Tyr Pro Glu Leu Ile Ser Arg Pro Ser Pro Pro
 515 520 525
 Thr Phe His Arg Phe Leu Pro Asp Leu Pro Pro Ser Arg Ser Ala Val
 530 535 540
 Glu Ile Ala Pro Thr Gln Val Thr Glu Thr Asp Glu Cys Arg Leu Asn
 545 550 555 560
 Gln Asn Ile Cys Gly His Gly Gln Cys Val Pro Gly Pro Ser Asp Tyr
 565 570 575
 Ser Cys His Cys Asn Ala Gly Tyr Arg Ser His Pro Gln His Arg Tyr
 580 585 590
 Cys Val Asp Val Asn Glu Cys Glu Ala Glu Pro Cys Gly Pro Gly Lys
 595 600 605
 Gly Ile Cys Met Asn Thr Gly Gly Ser Tyr Asn Cys His Cys Asn Arg
 610 615 620

Gly	Tyr	Arg	Leu	His	Val	Gly	Ala	Gly	Gly	Arg	Ser	Cys	Val	Asp	Leu	625	630	635	640
Asn	Glu	Cys	Ala	Lys	Pro	His	Leu	Cys	Gly	Asp	Gly	Gly	Phe	Cys	Ile	645	650	655	
Asn	Phe	Pro	Gly	His	Tyr	Lys	Cys	Asn	Cys	Tyr	Pro	Gly	Tyr	Arg	Leu	660	665	670	
Lys	Ala	Ser	Arg	Pro	Pro	Ile	Cys	Glu	Asp	Ile	Asp	Glu	Cys	Arg	Asp	675	680	685	
Pro	Ser	Thr	Cys	Pro	Asp	Gly	Lys	Cys	Glu	Asn	Lys	Pro	Gly	Ser	Phe	690	695	700	
Lys	Cys	Ile	Ala	Cys	Gln	Pro	Gly	Tyr	Arg	Ser	Gln	Gly	Gly	Gly	Ala	705	710	715	720
Cys	Arg	Asp	Val	Asn	Glu	Cys	Ser	Glu	Gly	Thr	Pro	Cys	Ser	Pro	Gly	725	730	735	
Trp	Cys	Glu	Lys	Leu	Pro	Gly	Ser	Tyr	Arg	Cys	Thr	Cys	Ala	Gln	Gly	740	745	750	
Ile	Arg	Thr	Arg	Thr	Gly	Arg	Leu	Ser	Cys	Ile	Asp	Val	Asp	Asp	Cys	755	760	765	
Glu	Ala	Gly	Lys	Val	Cys	Gln	Asp	Gly	Ile	Cys	Thr	Asn	Thr	Pro	Gly	770	775	780	
Ser	Phe	Gln	Cys	Gln	Cys	Leu	Ser	Gly	Tyr	His	Leu	Ser	Arg	Asp	Arg	785	790	795	800
Ser	Arg	Cys	Glu	Asp	Ile	Asp	Glu	Cys	Asp	Phe	Pro	Ala	Ala	Cys	Ile	805	810	815	
Gly	Gly	Asp	Cys	Ile	Asn	Thr	Asn	Gly	Ser	Tyr	Arg	Cys	Leu	Cys	Pro	820	825	830	
Leu	Gly	His	Arg	Leu	Val	Gly	Gly	Arg	Lys	Cys	Lys	Lys	Asp	Ile	Asp	835	840	845	
Glu	Cys	Ser	Gln	Asp	Pro	Gly	Leu	Cys	Leu	Pro	His	Ala	Cys	Glu	Asn	850	855	860	
Leu	Gln	Gly	Ser	Tyr	Val	Cys	Val	Cys	Asp	Glu	Gly	Phe	Thr	Leu	Thr	865	870	875	880
Gln	Asp	Gln	His	Gly	Cys	Glu	Glu	Val	Glu	Gln	Pro	His	His	Lys	Lys	885	890	895	
Glu	Cys	Tyr	Leu	Asn	Phe	Asp	Asp	Thr	Val	Phe	Cys	Asp	Ser	Val	Leu	900	905	910	
Ala	Thr	Asn	Val	Thr	Gln	Gln	Glu	Cys	Cys	Cys	Ser	Leu	Gly	Ala	Gly	915	920	925	
Trp	Gly	Asp	His	Cys	Glu	Ile	Tyr	Pro	Cys	Pro	Val	Tyr	Ser	Ser	Ala	930	935	940	
Glu	Phe	His	Ser	Leu	Val	Pro	Asp	Gly	Lys	Arg	Leu	His	Ser	Gly	Gln	945	950	955	960
Gln	His	Cys	Glu	Leu	Cys	Ile	Pro	Ala	His	Arg	Asp	Ile	Asp	Glu	Cys	965	970	975	
Ile	Leu	Phe	Gly	Ala	Glu	Ile	Cys	Lys	Glu	Gly	Lys	Cys	Val	Asn	Ser	980	985	990	
Gln	Pro	Gly	Tyr	Glu	Cys	Tyr	Cys	Lys	Gln	Gly	Phe	Tyr	Tyr	Asp	Gly	995	1000	1005	
Asn	Leu	Leu	Glu	Cys	Val	Asp	Val	Asp	Glu	Cys	Leu	Asp	Glu	Ser	Asn	1010	1015	1020	
Cys	Arg	Asn	Gly	Val	Cys	Glu	Asn	Thr	Trp	Arg	Leu	Pro	Cys	Ala	Cys	1025	1030	1035	1040
Thr	Pro	Pro	Ala	Glu	Tyr	Ser	Pro	Ala	Gln	Ala	Gln	Cys	Leu	Ser	Pro	1045	1050	1055	
Glu	Glu	Met	Glu	His	Ala	Pro	Glu	Arg	Arg	Glu	Val	Cys	Trp	Gly	Gln	1060	1065	1070	
Arg	Gly	Glu	Asp	Gly	Met	Cys	Met	Gly	Pro	Leu	Ala	Gly	Pro	Ala	Leu	1075	1080	1085	

Thr Phe Asp Asp Cys Cys Cys Arg Gln Pro Arg Leu Gly Tyr Gln Cys
 1090 1095 1100
 Arg Pro Cys Pro Pro Arg Gly Thr Gly Ser Gln Cys Pro Thr Ser Gln
 1105 1110 1115 1120
 Ser Glu Ser Asn Ser Phe Trp Asp Thr Ser Pro Leu Leu Leu Gly Lys
 1125 1130 1135
 Ser Pro Arg Asp Glu Asp Ser Ser Glu Glu Asp Ser Asp Glu Cys Arg
 1140 1145 1150
 Cys Val Ser Gly Pro Cys Val Pro Arg Pro Gly Gly Ala Val Cys Glu
 1155 1160 1165
 Cys Pro Gly Gly Phe Gln Leu Asp Ala Ser Arg Ala Arg Cys Val Asp
 1170 1175 1180
 Ile Asp Glu Cys Arg Glu Leu Asn Gln Arg Gly Leu Leu Cys Lys Ser
 1185 1190 1195 1200
 Glu Arg Cys Val Asn Thr Ser Gly Ser Phe Arg Cys Val Cys Lys Ala
 1205 1210 1215
 Gly Phe Thr Arg Ser Arg Pro His Gly Pro Ala Cys Leu Ser Ala Ala
 1220 1225 1230
 Ala Asp Asp Ala Ala Ile Ala His Thr Ser Val Ile Asp His Arg Gly
 1235 1240 1245
 Tyr Phe His
 1250

<210> 86

<211> 353

<212> PRT

<213> Homo sapiens

<400> 86

Met Ala Asn Asp Gln Lys Asn Ser Glu Ser Phe Pro Ala Lys Glu Asp
 1 5 10 15
 His Lys Lys Asp Asp Ala Ala Ala Pro Ala Glu Val Asp His Lys Asp
 20 25 30
 Glu Phe Ser Ala Ser Gln Pro His Pro Val Glu Asn Ile Val Leu Val
 35 40 45
 Gly Arg Thr Gly Asn Gly Lys Ser Ala Thr Gly Asn Ser Ile Val Arg
 50 55 60
 Ser Lys Val Phe Lys Ser Lys Thr Lys Ser Ser Gly Val Thr Met Glu
 65 70 75 80
 Cys His Ala Val Lys Ala Val Thr Pro Glu Gly Pro Ile Leu Asn Val
 85 90 95
 Ile Asp Thr Pro Gly Leu Phe Asp Leu Ser Val Ser Ala Glu Phe Ile
 100 105 110
 Gly Lys Glu Ile Val Lys Cys Leu Thr Leu Ala Asp Gly Gly Leu His
 115 120 125
 Ala Val Leu Leu Val Leu Ser Val Arg Thr Arg Ile Ser Gln Glu Glu
 130 135 140
 Glu Met Val Leu Ser Thr Leu Gln Val Leu Phe Gly Ser Lys Ile Val
 145 150 155 160
 Asp Tyr Leu Ile Val Val Phe Thr Gly Gly Asp Val Leu Glu Asp Asp
 165 170 175
 Gly Met Thr Leu Glu Asp Tyr Leu Gly Asp Asn Met Pro Asp Phe Leu
 180 185 190
 Lys Arg Val Leu Ile Leu Cys Gly Gln Arg Met Ile Leu Phe Asp Asn
 195 200 205
 Lys Thr Lys Asp Asp Glu Lys Lys Thr Lys Gln Val His Glu Leu Leu
 210 215 220

Lys Leu Ile Asp Leu Val Arg Lys Gln Asn Asn Asn Ile Pro Tyr Thr
 225 230 235 240
 Asp Glu Met Tyr His Met Ile Lys Glu Glu Asn Glu Arg His Lys Lys
 245 250 255
 Glu Gln Glu Glu Leu Glu Ser Lys Gly His Ser Glu Glu Gln Leu Ala
 260 265 270
 Ala Leu Met Lys Glu Leu Gln Ile Met Asn Glu Arg Asn Leu Lys Ala
 275 280 285
 Met Ala Glu Met Met Glu Lys Asn Met Lys Ile Ala Met Glu Ala Gln
 290 295 300
 Glu Lys Leu Phe Glu Gln Arg Glu Lys Ala Gln Glu Met Ser Tyr Gln
 305 310 315 320
 Gln Lys Met Glu Met Gln Glu Lys Leu Lys Gln Met Glu Gly Arg Met
 325 330 335
 Arg Ala Glu Met Glu Ala Gln Met Leu Ser Arg Gln Gln Cys Ser Ile
 340 345 350
 Leu

<210> 87

<211> 4317

<212> DNA

<213> Homo sapiens

<400> 87

```

cctcctgctg tccccctccct acccttggtc tctcgccccg ctctgccctc tgctaccaac      60
actcgatccc ctgctcgggc tgcacctcca atctccgagg gtcgtgcggc cccgatgccc      120
cgggccccga gcggtgcccc cggcctggcc cctgcgatgc gccaggccgg cggattgggg      180
ctgctggcac tactcctgct ggcgctgctg ggccccggcg gccagggggt gggccggccg      240
ggcagcgggg cacaggcggg ggcggggcgc tgggccccaa gcttcaagggt ggtctttgcg      300
cctgtgatct gcaagcggac ctgtctgaag ggccagtgtc gggacagctg tcagcagggc      360
tccaacatga cgtcatcggg agagaacggc cacagcaccg acacgctcac cggttctgcc      420
ttccgcgtgg tgggtgtgcc tctacctgac atgaacggtg gccagtgtc ttcccgaaac      480
cagtgcctgt gtcccccgga tttcacgggg cgcttctgcc aggtgcctgc tgcaggaaac      540
ggagctggca cggggagttc agggccccgg tggcccgacc gggccatgtc cacaggcccc      600
ctgccgcccc ttgccccaga aggagagtct gtggctagca aacacgccat ttacgcggtg      660
cagggtgatcg cagatcctcc cgggccccgg gagggtcctc ctgcacaaca tgcagccttc      720
ttggtgcccc tggggccagg acaaattctc gcagaagtgc aggtcccgcc ccccggtggtg      780
aacgtgcgtg tccatcacc tctgaagct tccgttcagg tgcaccgat cgagggggccg      840
aacgctgaag gccagcctc tcccagcac ttgctgcgc atcccagcc cccgcaccgc      900
aggccaccca ctcaaaagcc actgggcccgc tgcttccagg acacattgcc caagcagcct      960
tgtggcagca accctttgcc tggccttacc aagcaggaag attgctgcgg tagcatcggt      1020
actgcctggg gacaaagcaa gtgtcacaa gtcccacagc ttcagtatac aggggtgcag      1080
aagcctgtac ctgtacgtgg ggaggtgggt gctgactgcc cccagggtca caagaggctc      1140
aacagcacc actgccagga tatcaacgaa tgtgcgatgc ccgggaatgt gtgcatggt      1200
gactgcctca acaacctggt ctcttatcgc tgtgtctgcc cgcccggtca tagcttgggt      1260
ccctcgcag cacagtgcac tgccgacaaa ccagaggaga agagcctgtg tttccgcctt      1320
gtgagcaccg aacaccagt ccagcaccct ctgaccacac gctaaccgg ccagctctgc      1380
tgctgtagtg tgggtaaagc ctgggggtgcc cggtgccagc gctgcccggc agatggtaca      1440
gcagccttca aggagatctg ccccggtggt gaaagggtag catatcctca cctcccacca      1500
gacgtcacc atccagggg aaagcagctt ctccctcttc ctgccccga cgggccacc      1560
aaaccccagc agcttcctga aagccccagc ctagcaccac ccctcgagga cacagaggaa      1620
gagagaggag tgaccatgga tccaccagt agtgaggagc gatcggtgca gcagagccac      1680
cccactacca ccacctcacc ccccggcct taccagagc tcatctctcg cccctcccca      1740
cctaccttcc accggttcc tccagacttg ccccatccc gaagtgcagt ggagatcgcc      1800
cccactcagg tcacagagac cgatgagtgc cgattgaacc agaatatctg tggccatgga      1860
cagtgtgtgc ctggccctc ggattactcc tgccactgca acgctggcta ccggtcacac      1920
ccgcagcacc gctactgtgt tgatgtgaac gagtgcgagg cagagccctg cggccccggg      1980

```

```

aaaggcatct gtatgaacac tgggtggctcc tacaattgtc actgcaaccg aggctaccgc 2040
ctccacgtgg gtgcaggggg cgcctcgtgc gtggacctga acgagtgcgc caagcctcac 2100
ctgtgtgggg acggtggctt ctgcatcaac ttccctgggtc actacaaatg caactgctat 2160
cctggctacc ggctcaaggc ctcccgaaccg cccatttgcg aagacatcga cgagtgtcgc 2220
gaccctagca cctgccctga tggcaaatgt gaaaacaaac ctggcagctt caagtgcac 2280
gcctgccagc ctggctaccg tagccagggg ggcggggcct gtcgtgatgt caacgaatgc 2340
tccgagggta cccctgctc tcctggatgg tgtgagaacc ttccgggttc ttaccgttgc 2400
acgtgtgccc aggggatacg aaccgcaca ggacgcctca gttgcataga cgtggatgag 2460
tgtgaggtcg ggaaagtgtg ccaagatggc atctgcacga acacaccagg ctctttccag 2520
tgtcagtgcc tctccggcta tcattctgtc agggatcgga gccgctgtga ggacattgat 2580
gaatgtgact tccctgcggc ctgcatcggg ggtgactgca tcaataccaa tggttcctac 2640
agatgtctct gtcccctggg tcatecgggtg gtgggcggca ggaagtgcaa gaaagatata 2700
gagagtgcga gccaggaccc aggcctgtgc ctgccccatg cctgcgagaa cctccagggc 2760
tcctatgtct gtgtctgtga tgagggtttc acactcacc aggaccagca tgggtgtgag 2820
gaggtggagc agccccacca caagaaggag tgctacctta acttcgatga cacagtgttc 2880
tgtgacagcg tattggctac caatgtcact cagcaggaat gctgttgctc tctgggagct 2940
ggctggggag accattgcga aatctatccc tgtccagtct acagctcagc cgaatttcac 3000
agcctgggtg ctgatgggaa aaggctacac tcaggacaac aacattgtga actatgcatt 3060
cctgcccacc gtgacatcga cgaatgcata ttgtttgggg cagagatctg caaggagggc 3120
aagtgtgtga acacgcagcc cggctacgag tgctactgca agcagggtct ctactacgat 3180
ggcaacctgc tggagtgcgt ggacgtggat gagtgtcttg atgagtctaa ctgcaggaac 3240
ggagtgtgtg agaacacacg tggcggctac cgctgtgcct gcactccgcc ggacaggtac 3300
agtcccgcac agggccagt tctgatcccg gagagatgga gcacgcccc gagagacgtg 3360
aagtgtgtcg gggccagcga ggagaggacg gcatgtgtat gggggccctg ggcgggacct 3420
gccctcactt ttgatgactg ctgctgccgc cagccgcggc tgggtacca gtgcagaccg 3480
tgcccgccac gtggcaccgg gtcccagtgc ccgacttcac agagtgcgag caattctttc 3540
tgggacacaa gccccctgct actggggaag tctccgcgag acgaagacag ctgagaggag 3600
gattcagatg agtgccgttg tgtgagcggg cgctgtgtgc cacggccagg cggggcggtg 3660
tgcgagtgtc ctggaggctt tcagctggac gcctcccgtg cccgctgcgt ggacattgat 3720
gagtgccgag aactgaacca gcggggactg ctgtgtaaga gcgagcgggtg cgtgaacacc 3780
agtggatcct tccgctgtgt ctgcaaagct ggcttcacgc gcagccgcc tcacgggcct 3840
cgctgcctca gcgcgcgcgc tgatgatgca gccatagccc acacctcagt gatcgcacat 3900
cgaggggtatt ttcactgaaa gtggagacag acaagtacat cctttgtctc tgaccaaaccg 3960
agagcatgga cccaaggatc cttcaggggc cacaatctc cttccacac cccaaaccca 4020
aggtgtctct gtctgcagag tgctgtctgc tttctccaa ggggtgattc tagaaacttc 4080
gacatcagat ctgccccttt aatttactct tggctttcaa ggcaaattga tattcacatc 4140
caaagcgggc agcatcaact gcttggcggg ttggactgag ctgggaccca ggatgtgaaa 4200
tagaatttat tgtggctctg attatgtaca ctagatgtgc ctgacctgct gaccaggctc 4260
acatggtttg tacaataaat acatccgccg ggaaaaaaaa aaaaaaaaaa aaaaaaa 4317

```

<210> 88

<211> 299

<212> PRT

<213> Mus musculus

<400> 88

```

Ser Val Ser Trp Pro Asp Thr Ala Gln Gly Thr Met Ala Asn Leu Ile
 1           5           10          15
Leu Thr Ala Leu Glu Lys Ala Thr Leu Phe Leu Glu Asp Arg Leu Pro
      20           25           30
Thr Ile Asn Leu Asp Gly Val Val Gly Phe Gln Val Leu Glu Val Gln
      35           40           45
Leu Arg Gly Val Gln Glu Lys Trp Ala His Lys Pro Leu Leu Gln Pro
      50           55           60
Leu Ser Met Arg Ala Gly Gln Met Ala Asn Thr Leu Ser Ala Leu Leu
      65           70           75           80

```


Gln Lys Ser Ile Phe Tyr Leu Lys Gln Ser Asp Pro Thr Tyr Leu Arg
 85 90 95
 Glu Phe Gln Pro Ser Ile Gln Pro Gly Phe Trp Lys Leu Pro Asn Asp
 100 105 110
 Trp Thr Arg Thr Asn Ala Ser Leu Val Tyr Pro Trp Leu Glu Pro Leu
 115 120 125
 Asp Ser Phe Ser Glu Glu Ser Ser Asp Val Cys Leu Val Gln Leu Leu
 130 135 140
 Gly Thr Gly Thr Asp Ser Ser Gln Pro Cys Arg Leu Ser Asn Phe Cys
 145 150 155 160
 Arg Thr Leu Met Thr Lys Ala Gly Cys Ser Gly Tyr Ser Leu Ser His
 165 170 175
 Gln Leu Leu Phe Phe Leu Trp Ala Arg Met Gln Gly Cys Thr Glu Gly
 180 185 190
 Leu Phe Leu Gln Ser Gln His Tyr Met Asp Ile Phe Cys Ala Asn Met
 195 200 205
 Met Glu Leu Asn His Arg Ala Glu Ala Val Gly Tyr Ala Tyr Pro Thr
 210 215 220
 Gln Asp Leu Phe Met Glu Asn Ile Met Phe Cys Gly Met Ala Gly Phe
 225 230 235 240
 Ser Asp Phe Tyr Lys Leu Arg Trp Leu Glu Ala Ile Leu Ser Trp Gln
 245 250 255
 Asn Pro Gln Val Gly Cys Phe Gly Arg Pro Asp Thr Lys Gly Glu Pro
 260 265 270
 Ser Glu Val Pro His Gln Gln Gly Ile Leu Arg Arg Val Arg Arg Arg
 275 280 285
 Glu Lys Leu Phe Ala Asp Gly Cys Ser Cys His
 290 295

<210> 89

<211> 17

<212> PRT

<213> Mus musculus

<400> 89

Asn Thr Ala Thr Ala Val Ala Ala Leu Gly Gly Phe Leu Tyr Ile Leu
 1 5 10 15
 Ala

<210> 90

<211> 50

<212> PRT

<213> Mus musculus

<400> 90

Glu Tyr His Pro Asp Asn Gly Asp Ala His Pro Glu Tyr Tyr Pro Asn
 1 5 10 15
 His Gly Asp Pro Tyr Ser Ser Ser Gln Ser Pro Ala Ser Asn Tyr Gln
 20 25 30
 Asp Gly Ala Ala Gly Pro Asp Val Gln Arg Thr Gly Arg Pro Leu Ser
 35 40 45
 Val Ser
 50

<210> 91

<211> 199

<212> PRT

<213> Mus musculus

<400> 91

```

Met Leu Leu Leu Leu Leu Gly Phe Leu Gly Pro Ala Ala Cys Trp Ala
 1           5           10           15
Leu Gly Pro Ala Gly Pro Gly Ser Ser Glu Leu Arg Ser Ala Phe Ser
      20           25           30
Ala Ala Arg Thr Thr Pro Leu Glu Gly Thr Ser Glu Met Ala Val Thr
      35           40           45
Phe Asp Lys Val Tyr Val Asn Ile Gly Gly Asp Phe Asp Ala Ala Thr
      50           55           60
Gly Arg Phe Arg Cys Arg Val Pro Gly Ala Tyr Phe Phe Ser Phe Thr
      65           70           75           80
Ala Gly Lys Ala Pro His Lys Asn Leu Ser Val Met Leu Val Arg Asn
      85           90           95
Arg Asp Glu Val Gln Ala Leu Ala Phe Asp Lys Gln Arg Arg Pro Gly
      100          105          110
Ala Arg Arg Ala Ala Ser Gln Ser Ala Met Leu Gln Leu Asp Tyr Gly
      115          120          125
Asp Thr Val Trp Leu Arg Leu His Gly Ala Pro His Tyr Ala Leu Gly
      130          135          140
Ala Pro Gly Ala Thr Phe Ser Gly Tyr Leu Val Tyr Ala Asp Ala Asp
      145          150          155          160
Ala Asp Ala Pro Ala Arg Gly Pro Ala Ala Pro Glu Pro Arg Ser Ala
      165          170          175
Phe Ser Ala Arg His Ala Thr Trp Trp Ala Pro Asn Pro Pro Arg Pro
      180          185          190
Ala Pro Arg Arg Leu Ala Phe
      195

```

<210> 92

<211> 597

<212> DNA

<213> Mus musculus

<400> 92

```

atgctgctgc tcttgctggg cttcctaggc ccggcggcct gctgggcact gggcccggct      60
ggccctggct cctcggagct gcggtcagcc ttctcggcgg ctcgcaccac cccgctggag      120
ggcacgtcgg agatggcggt gaccttcgac aaggtgtacg tgaacatcgg ggggtgacttc      180
gacgcagcca ccgggcggtt ccgctgtcgc gtgccgggcg cctacttctt ctcttcacg      240
gccggcaagg ccccgcacaa aaacctgtcg gtgatgctgg tgcgcaaccg cgacgaagtg      300
caggcgctgg ctttcgacaa gcagcgacgg ccaggcgcg cgcgcgcggc cagccaaagc      360
gccatgctgc agctcgacta cggcgacacg gtgtggctgc ggctgcacgg cgctccgcat      420
tacgcgctcg gcgcgccggg cgccaccttc agcggctacc tgggtgtacg ggacgccgac      480
gccgacgcgc ctgcgcgcgg gcccgcgggc ccggagccgc gctcggcctt ctccgcgcgc      540
cacgccacct ggtgggctcc gaaccgcgcc cggccgcgc cacggcgttt ggccttc      597

```

<210> 93

<211> 126

<212> PRT

<213> Mus musculus

<400> 93

```

Ala Phe Ser Ala Ala Arg Thr Thr Pro Leu Glu Gly Thr Ser Glu Met
 1           5           10           15
Ala Val Thr Phe Asp Lys Val Tyr Val Asn Ile Gly Gly Asp Phe Asp
      20           25           30

```

Ala	Ala	Thr	Gly	Arg	Phe	Arg	Cys	Arg	Val	Pro	Gly	Ala	Tyr	Phe	Phe
		35					40					45			
Ser	Phe	Thr	Ala	Gly	Lys	Ala	Pro	His	Lys	Asn	Leu	Ser	Val	Met	Leu
	50					55					60				
Val	Arg	Asn	Arg	Asp	Glu	Val	Gln	Ala	Leu	Ala	Phe	Asp	Lys	Gln	Arg
65					70					75					80
Arg	Pro	Gly	Ala	Arg	Arg	Ala	Ala	Ser	Gln	Ser	Ala	Met	Leu	Gln	Leu
				85					90					95	
Asp	Tyr	Gly	Asp	Thr	Val	Trp	Leu	Arg	Leu	His	Gly	Ala	Pro	His	Tyr
			100					105					110		
Ala	Leu	Gly	Ala	Pro	Gly	Ala	Thr	Phe	Ser	Gly	Tyr	Leu	Val		
		115					120					125			